Met Office

Ensemble Exploitation Strategic Action Work Package 1

Extracting untapped information from ensembles in support of advice, guidance and storylines

Forecasting Research Technical Report No. 664

November 2024

Authors:

Helen Titley (EE WP1 co-lead), Steve Willington (EE WP1 co-lead), Brian Ancell (Texas Tech University), Daniel Etheridge, Rebekah Hicks, Caroline Jones, Robert Neal, Phil Relton, Nigel Roberts (former WP1 co-lead, now at ECMWF), Adrian Semple and Brent Walker.

DOI: https://doi.org/10.62998/lduw8943

© Crown Copyright 2024, Met Office

Document history

Version	Purpose	Date
0.1	Draft	30th September 2024
1.0	Near complete draft shared with Rosa	23rd October 2024
1.1	Complete report available for review	8th November 2024
1.2	Final report	29th November 2024

Prepared by

WP1 project co-leads and WP1 project team

Reviewed by

Ken Mylne, Science Fellow in Exploitation of Ensembles, 13th November 2024

Rosa Barciela, Head of Ensembles Exploitation Strategic Action, 26th November 2024

Authorised for issue by

Rosa Barciela, Head of Ensembles Exploitation Strategic Action, 26th November 2024

© Crown copyright 2024, Met Office

Contents

1. Introduction....5
2. Ensemble Sensitivity Analysis and Ensemble Sub-setting....6
2.1. Ensemble Sensitivity Analysis....6
2.1.1. Background and introduction to ESA....6
2.1.2. ESA methodology....7
2.1.3. Description of the two ESA research prototype tools....8
2.1.4. Example case study using ESA: Uncertainty in the extra-tropical transition of Hurricane Kirk and its subsequent evolution as an extra-tropical cyclone 12
2.1.5. Key findings from the ESA trial....18
2.2. Ensemble Sub-setting (ESS)....22
2.2.1. Introduction to ESA-based Ensemble Sub-setting (ESS)....22
2.2.2. ESS methodology implemented in EE WP1....22
2.2.3. Description of the ESS prototype display including illustrative case 23
2.2.4. Objective evaluation....25
2.3. Recommendations for ESA and ESS....32
3. Difference tracking....34
3.1. Comparison Of Difference Tracking and Ensemble Sensitivity Analysis. 34
3.1.1. Introduction....34
3.1.2. Comparisons of DT and ESA....35
3.1.3. Conclusions....37
3.2. Feasibility Study of Difference Tracking Applied to Ensembles....38
3.2.1. Introduction....38
3.2.2. The Feasibility Study....38
3.2.3. Conclusions....40
3.3. Recommendations....41
4. Feature-based clustering....42
4.1. Background and introduction....42

© Crown copyright 2024, Met Office

4.2. Methodology....44
4.3. Example clustering output....47
4.4. Recommendations for future work....53
References....56

Glossary

CRPS	Continuous Ranked Probability Score
DT	Difference Tracking
ECMWF	European Centre for Medium-Range Weather Forecasts
EE	Ensemble Exploitation
EOF	Empirical Orthogonal Function
ESA	Ensemble Sensitivity Analysis
ESS	Ensemble sub-setting
EWH	Expert Weather Hub
FSS	Fractions Skill Score
GM	Global Model (Met Office UM global deterministic model)
GPH	Geopotential Height
MAE	Mean Absolute Error
MOGREPS-G	Global component of the Met Office Global and Regional Ensemble Prediction System
MSE	Mean square error of the ensemble mean
MSLP	Mean Sea Level Pressure
Op Mets	Operational Meteorologists
OTD	On the day (ESA)
PCA	Principal Component Analysis
PDF	Probability density function
PFJ	Polar Front Jet
PROJ	Projection method, as defined in Ancell (2016)
RM	Representative Member
STJ	Subtropical Jet
WP1	Work Package 1

© Crown copyright 2024, Met Office

1. Introduction

The Ensemble Exploitation (EE) Strategic Action was created to provide the dedicated effort required to embed ensembles at the heart of our products and services and allow us to capitalise on the insight ensembles provide for weather forecasting. The focus of EE Work Package 1 (WP1) in the final year of the strategic action has been on developing ways of connecting Expert Operational Meteorologists (Op Mets) with ensemble forecasts in a way that enables investigative interaction with the data rather than passive interaction with pre-determined products. A family of research prototype tools have been developed that provide novel ways to extract untapped information from ensemble forecasts to allow improved understanding of the forecast evolution and drivers of uncertainty to better inform storylines, advice, and guidance.

This report describes the three main areas that have been developed:

- Section 2 describes a family of tools based on Ensemble Sensitivity Analysis (ESA), a technique that computes the relationship between a forecast event of importance and the atmospheric state at times leading up to the event, giving an enhanced understanding of forecast evolution and synoptic drivers. Ensemble sub-setting is also explored, where early forecast errors in sensitive areas are used to derive a subset of the ensemble.
- Section 3 gives a feasibility study of how to apply difference tracking, a well-used tool in the Expert Weather Hub (EWH) for exploring differences in model evolution between different deterministic model runs, to ensemble forecasts.
- Section 4 describes a new method to cluster ensemble forecasts based on identifying features in each ensemble member and comparing the identified features across the ensemble to output clusters, from which representative members can also be identified.

Each of the methods has been initially trialled on the Global component of the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G). Each section in the report will summarise the development work that has been carried out, illustrate this with forecast examples, summarise user feedback, and give a series of future recommendations on how this work will be taken forward after the end of the strategic action.

© Crown copyright 2024, Met Office

2. Ensemble Sensitivity Analysis and Ensemble Sub-setting

Ensemble Sensitivity Analysis (ESA) is a technique that highlights the relationship between a chosen forecast event and the wider atmospheric state at times leading up to the event, giving an enhanced understanding of ensemble evolution and synoptic drivers.

Three research prototype tools have been developed to allow the Expert Weather Hub (EWH) to trial using ESA and Ensemble sub-setting (ESS) in their forecasting processes:

- The ***ESA "on the day"*** research prototype tool runs routinely in real time twice a day on the 00/12UTC runs of MOGREPS-G using a predefined set of responses and sensitivity parameters (as defined in Section 2.1).
- The ***ESA "on demand"*** research prototype is a more interactive tool, with a more expansive set of options, that allows Op Mets to select a specific response and sensitivity variable of their choice, to allow a new ESA run to be requested by the Op Met tailored to the specific weather event of interest that day.
- The ***Ensemble sub-setting*** (ESS) research prototype uses additional information (an updated analysis or satellite imagery) to see which ensemble members are drifting from what has already happened in those areas where the ESA has identified sensitivity, and updates the ensemble to exclude those members with the greatest errors in those areas.

The three prototypes have been developed and managed using a Met Office GitHub repository (https://github.com/MetOffice/ee-esa-dev), which stores the code and documentation, and contains a record of code changes and reviews. ESA code from Austin Coleman (former Texas Tech University PhD student) and ESS code from Brian Ancell (Texas Tech University) were ported and adapted extensively for use within the Met Office computing environments and data formats. Multiple new visualisations were added to aid analysis, interactive webpage interfaces were created, and optimised parallelisation implemented to dramatically speed up the code.

This section will describe the ESA and ESS methods, introduce the research prototype tools, illustrate how to use the tools and interpret the output using case studies, summarise feedback from the EWH trial, and give recommendations for the future.

© Crown copyright 2024, Met Office

2.1. Ensemble Sensitivity Analysis

2.1.1. Background and introduction to ESA

Ensemble Sensitivity Analysis (Ancell & Hakim, 2007) (Torn & Hakim, 2008) (Ancell & Coleman, 2022) is a statistical technique based on linear regression used to help understand the evolution of ensemble forecasts by computing the statistical relationship between some forecast aspect of importance (event) and the atmospheric state at lead times before, at, and after the event. While ensemble sensitivity is statistical in nature, it is linked to the dynamics of perturbation evolution and shows how changes in weather features would evolve to change a chosen response. Typical uses of ESA include sensitivity-based observation targeting (Ancell & Hakim, 2007), monitoring precursors to extreme weather events (Ancell & Coleman, 2022), and informing ensemble sub-setting to reduce forecast errors (Coleman & Ancell, 2020).

The forecast weather event of interest, which is represented by a chosen scalar function of a forecast variable in an area of interest at a time of interest, is defined as the “response” function e.g. the maximum wind speed over the UK. The atmospheric state in a much larger chosen domain is then represented by a forecast “sensitivity” variable e.g. the mean sea level pressure or the geopotential height at 300hPa over a North American-North Atlantic-European domain. Ensemble sensitivity can then be determined by calculating the regression coefficients (slopes of the linear regression) between the response function and the sensitivity variables for the ensemble at different forecast lead times (Ancell & Coleman, 2022). In this way sensitive areas can be identified for times leading up to a weather event, aiding understanding of what weather features are driving uncertainty in the ensemble forecast, how the ensemble is evolving, and whether there are any key points in the evolution that disproportionately lead to member diversity.

2.1.2. ESA methodology

The **response (R)** is the name given to a particular aspect of the weather event of interest. It is identified using the following:

i) A forecast variable e.g. wind speed at 600m *(response variable)*
ii) A scalar response function to measure a feature of that variable, e.g. number of grid points > 20 m/s *(response function)*
iii) An area covering the relevant geography e.g. a square covering the UK *(response area)*
iv) The validity time or period of the event *(response time)*

The scalar function is then extracted using the variable and area, for each ensemble member, to give the **response (R).** If there are 36 members in your

© Crown copyright 2024, Met Office

ensemble, operating on their data with the response function will produce 36 response values. These response values are then taken forward into regression coefficient calculations to obtain the sensitivities.

Ensemble sensitivity is then calculated over a larger domain for an atmospheric state (sensitivity) variable (x), chosen due to an expectation of some correlation with the response. Raw sensitivity is the regression coefficient, or slope of the linear regression, calculated between the sensitivity variable at the grid point (x) and the scalar response value (y) for each ensemble member.

This sensitivity is converted to "standardised sensitivity" by multiplying by the ensemble standard deviation of the sensitivity variable, to highlight where the members differ as well as being correlated. This results in the units becoming the same as those of the response variable. Conceptually, standardised sensitivity represents the change to the response due to expected ensemble perturbations, unlike the raw sensitivity, which estimates changes to the response for arbitrary perturbations to the sensitivity variable. The equation for standardised sensitivity can be written as follows:

$$Standardised\,Sensitivity = \left[\frac{covariance(x,R)}{variance(x)}\right] Stdev(x)$$

The resulting standardised sensitivity field is plotted to allow visual identification of areas with direct and indirect dynamical links to the weather event.

2.1.3. Description of the two ESA research prototype tools

ESA on the day

The ESA "on the day" research prototype tool runs routinely in real time twice a day on the 00/12UTC runs of MOGREPS-G using a predefined set of responses and sensitivity parameters. The options available in ESA on the day are shown in Table 2.1.

Response variables and functions	• Wind speed at 600 metres; 3 response functions are available: - number of points in the response area >20m/s in each member - Principal Component Analysis (PCA): a technique to find the leading modes of variability (principal components) in the data that capture the most variance; the function values show how closely each member aligns with that component. - 95th percentile across the response area in each ensemble member • Screen level temperature; 2 response functions are available: - mean temperature across the response area in each

© Crown copyright 2024, Met Office

	ensemble member - 95th percentile across the response area in each ensemble member • Accumulated total precipitation; 2 response functions are available as follows: - 6-hour maximum precipitation accumulation across the response area in each member. - 24-hour maximum precipitation accumulation across the response area in each member.
Response times	T+12 to T+156 at 12-hour intervals
Response area	A polygon covering the UK
Sensitivity variables	• Geopotential height at 300hPa • Geopotential height at 500hPa • Mean sea level pressure • Wind speed at 300hPa • Water vapor pseudo imagery

Table 2.1. The options available in ESA on the day.

A screenshot of the ESA on the day web interface is shown in Figure 2.1, and shows how the pre-created response and sensitivity images can be selected by the user using a series of drop-down menus. Figure 2.2 then displays larger images of the key response and sensitivity plots along with descriptions of what they show.

© Crown copyright 2024, Met Office

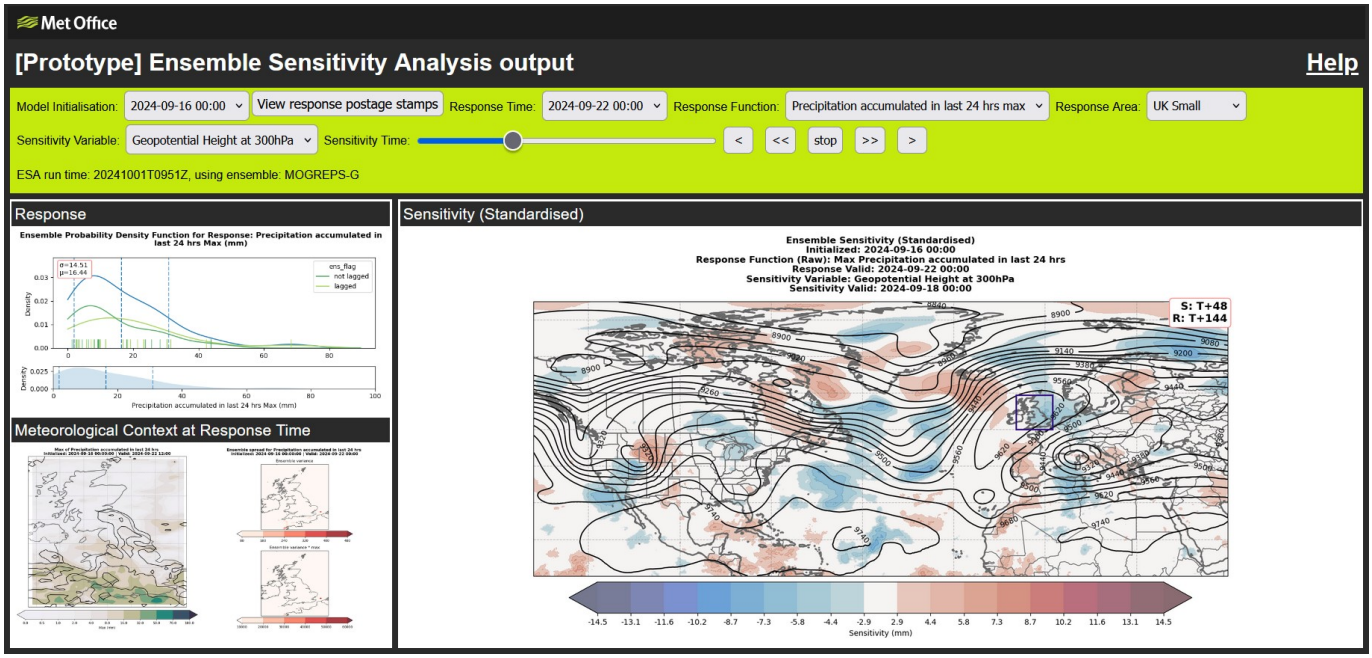
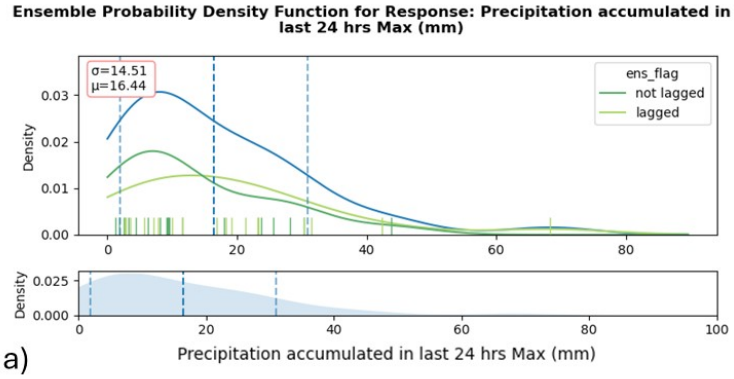


Figure 2.1. ESA on the day user interface.

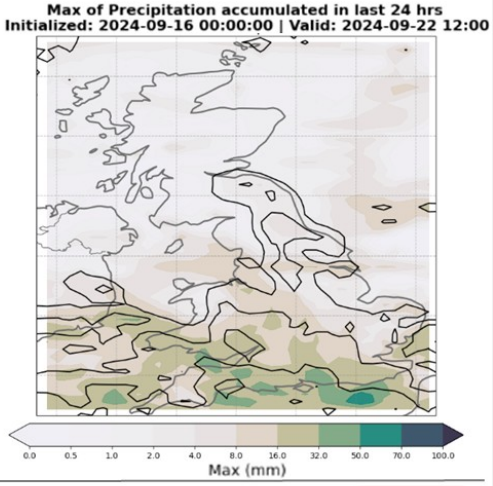
© Crown copyright 2024, Met Office



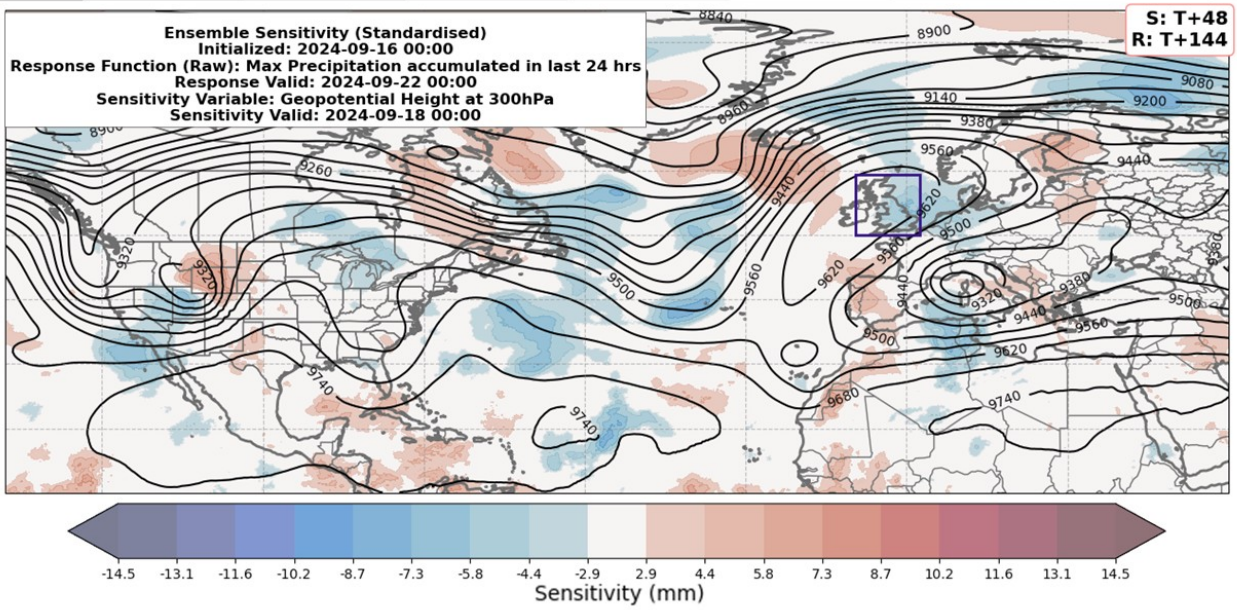
a)

Probability Density Function (PDF) of response: The blue line shows the distribution of the response values - in this case the maximum 24-hour (T+120-T+144) precipitation accumulation in the response area - from each of the 36 ensemble members. The mean and standard deviation are shown by dashed lines. Each member value is shown by a green vertical line: dark green for the 18 latest members and light green for the 18 lagged members. The PDFs of the latest and lagged runs are also overlain. The smaller subplot at the bottom shows the PDF on a fixed axis to help when comparing between different response times and model runs.

Meteorological context plots: To give an idea of what the weather is like at the response time, the weather variable associated with the response function is also available to view in map form. For a threshold response function, the plot of contextual weather is a plot of the probability of the weather being above/below that threshold within the response area. For the aggregate type of response functions (e.g. max, mean, percentile), the same function is applied to produce the plot, but across the ensemble members at each point (rather than applying it for each ensemble member across the response area, as is the case when obtaining the response values). In the example shown, the meteorological context plot helps to show where the heaviest rainfall, which has been taken forward into the ESA calculation, is likely to be. Additional small plots (not shown) show the ensemble variance for the response variable on the top, and the variance multiplied by the maximum ensemble value on the bottom, to help highlight the areas of both extreme weather and uncertainty.



b)



c)

Sensitivity plot: Visualises the standardised sensitivity calculated at each point (the regression coefficients multiplied by the standard deviation of the sensitivity variable, as explained in section 2.1.2). In this case it shows the sensitivity of the response (T+120-T+144 maximum 24-hour precipitation across the area shown by the blue box in each member) to the geopotential height at 300hPa at T+48. The black contours show the ensemble mean for the sensitivity variable. The sensitivity values are displayed in a diverging blue to red colour map, with the scale auto-selected using the standard deviation of the response. If there is an area of negative sensitivity (blue in the plots), it indicates that a decrease in the sensitivity variable in this area results in an increase in the response values, while an area of positive sensitivity (red in the plots), indicates that an increase in the sensitivity variable in this area leads to an increase in the response values.

Figure 2.2. Example ESA on the day response and sensitivity plots, along with plot descriptions: a) Probability Density Function plot for the response; b) Meteorological context plot; c) the Sensitivity

© Crown copyright 2024, Met Office

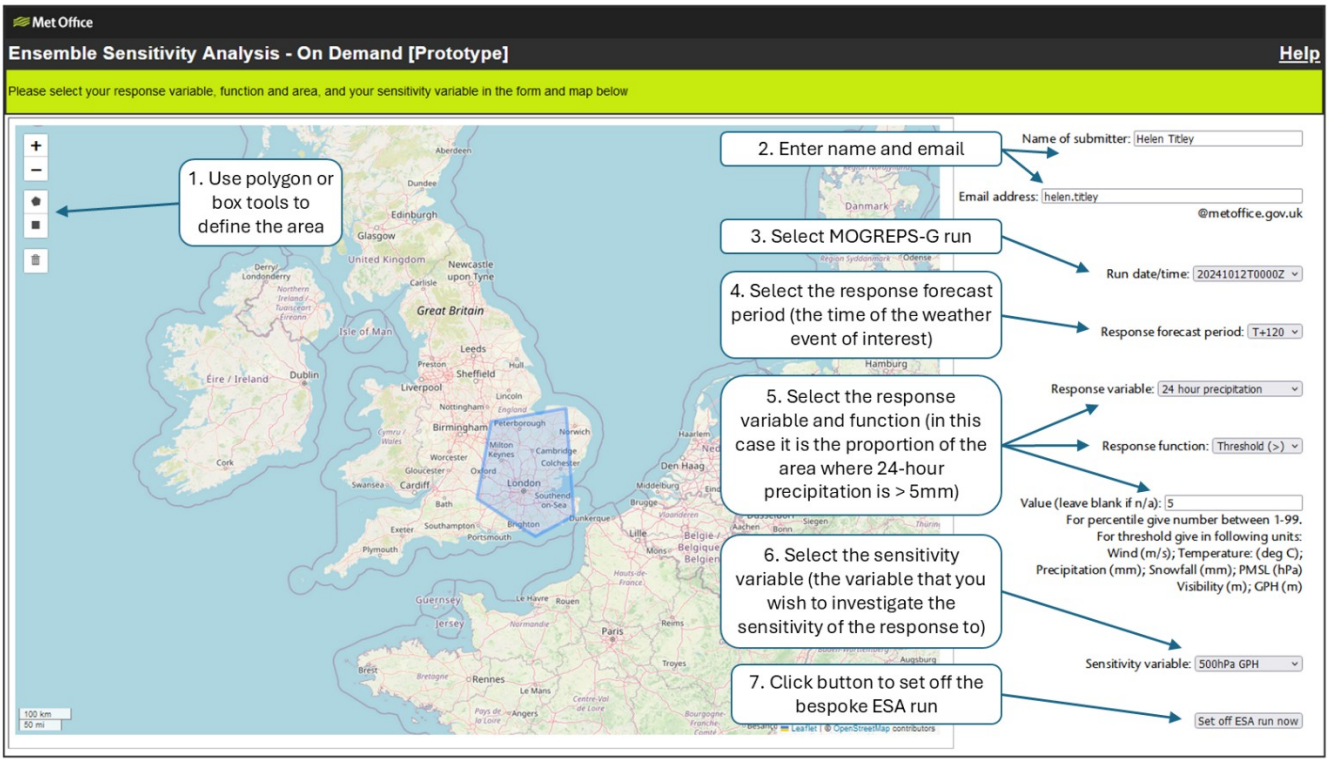
plot.

ESA on demand

The ESA “on demand” research prototype is a more interactive tool that allows Op Mets to select a specific response (area, variable and function) and sensitivity variable of their choice, to trigger a new ESA calculation tailored to the specific weather event of interest that day. It offers a larger number of variable and function options providing greater flexibility at the expense of a small delay in obtain the results.

The interface to submit a new ESA on demand calculation is shown in Figure 2.3. Any response area within the ESA domain can be selected via an interactive polygon. All the response variables that are available in the ESA on the day tool are available here, with additional variables such as snowfall, mean sea level pressure, and visibility, also being available, and the response functions can be tailored as required. Additional sensitivity variables are also available, such as 850 hPa wet bulb potential temperature. A new ESA calculation takes just a couple of minutes to complete, and when it is finished the outputs are displayed in a similar interface to the “ESA on the day” web page.

An analysis showed that 80 bespoke “on demand” submissions were received between early August and mid-September, by 12 different Op Mets in the EWH. Many different response and sensitivity combinations were chosen related to the weather of interest that day.

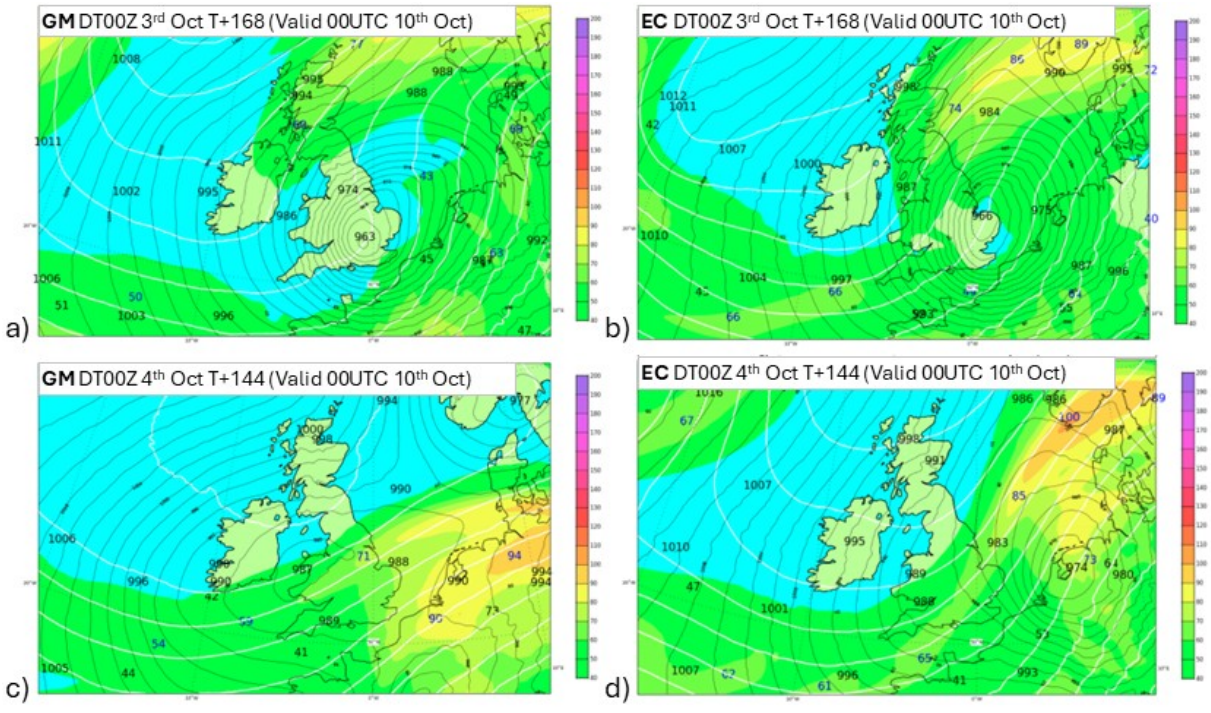


© Crown copyright 2024, Met Office

Figure 2.3. ESA on demand interface with annotations explaining the steps needed to set off a bespoke ESA calculation.

2.1.4. Example case study using ESA: Uncertainty in the extra-tropical transition of Hurricane Kirk and its subsequent evolution as an extra-tropical cyclone

This subsection provides a meteorological case-study example of the extratropical transition of Hurricane Kirk to show what ESA can reveal. The forecasting of the extra-tropical transition of tropical cyclones is a challenging process for NWP and humans alike. The evolution of Hurricane Kirk as it moved from the tropics and then underwent extra-tropical transition and eventual entrainment in the mid-latitude flow was no different and on 3rd October 2024 the Expert Weather Hub (EWH) was presented with a solution from the DT00Z 3rd October Global Model (GM) run (Figure 2.4a), with support from other centres such as ECMWF (Figure 2.4b), for the remnants of 'Kirk' to move across the south of the UK as a worryingly deep depression. Subsequent forecasts of the GM moved away from a deep depression and while ECMWF followed the same trend it did so to a lesser degree and continued to retain the threat of a deep depression (e.g. Figure 2.4c). To understand why this was happening in the GM it is necessary to understand the drivers in the early model conditions and how the different evolutions subsequently evolve, and we can use the ESA tools to help reveal these. What follows is an example of the type of information which can be extracted and how it can help understanding and hence decision making by the EWH.



© Crown copyright 2024, Met Office

Figure 2.4: Forecasts verifying 00Z 10th October 2024, from DT00Z 3rd October 2024 GM (a) and ECMWF Det (b), and DT00Z 4th October 2024 GM (c) and ECMWF Det (d). Black contours show Mean Sea Level Pressure (MSLP) and the colours show 250hPa windspeed.

GM output from DT00Z 3rd October drew ‘Kirk’ into the mid-latitude flow in a preferential area lying between the left exit of the Subtropical Jet (STJ) and right entrance of the Polar Front Jet (PFJ) to its north. This preferential location and phasing allowed ‘Kirk’ to deepen rapidly into a potentially impactful storm as it was forecast to cross the south of the United Kingdom during the afternoon and night of 9th October before continuing into the North Sea during 10th October. By the morning of the 4th October, more recent runs of the GM, the 12Z GM on 3rd October and 00Z GM 4th October, were trending southwards towards a lower latitude shallow feature presenting the EWH with a challenge regarding guidance and advice.

These later runs of the GM, for instance the DT00Z 4th October run, located ‘Kirk’ west and then northwest of the earlier DT00Z 3rd October run during the stage of extra-tropical transition between 7th and 8th October (Figure 2.5). This allows a stronger upper ridge to build ahead of ‘Kirk’ in the subtropical flow, which ‘Kirk’ then must pass through. By the time it has done this the trough forward is relaxing as the northwesterly PFJ extends more deeply into Iberia. The outcome is that ‘Kirk’ is steered east as a filling feature across the north of Iberia and then into central France. Note also the persistence of the pre-existing depression over the UK in the more recent runs also helps to prevent ‘Kirk’ being taken northeast.

The track comparison in Figure 2.5 shows that ‘Kirk’ at 00Z 8th October 2024 is a bit further north and west during its extratropical transition in the GM run of 4th October when compared to the 00Z run from the 3rd October. The fact that the forecast with the storm further north at this earlier stage ends up with it further south in the end is perhaps counter-intuitive and highlights that the dynamics are complicated, i.e. that a simple linear continuation of an earlier positional difference does not lead to the correct outcome.

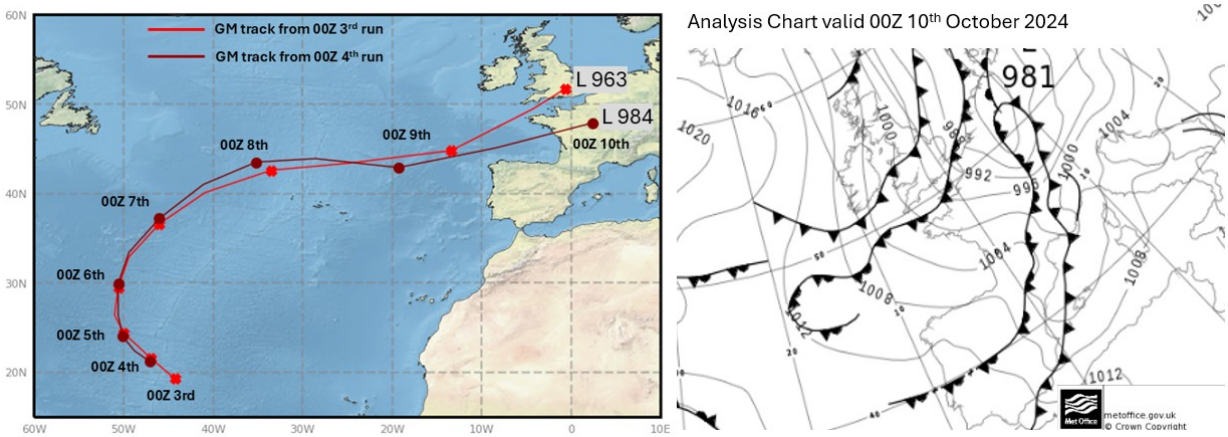


Figure 2.5. Forecast track of ‘Kirk’. Red GM DT00Z 3rd October, Maroon GM DT00Z 4th October. Markers show the 00Z locations. The GM DT00Z 4th October run proved to be considerably nearer

to the actual evolution although it was approximately 6 hours slow by the end of the forecast period (see Met Office Analysis Chart valid 00Z 10th October (right)).

The ESA sensitivity plot in Figure 2.6a is showing that the members with the strongest wind over the UK at the later time, later on 9th October and during 10th October, had lower surface pressure ahead of the storm at the earlier time (blue to the east of the depression), but also less troughing to the south of the depression to the west of Ireland (pink = higher MSLP at earlier time = windier over UK at later time). In other words, a faster eastward progression of 'Kirk' during its extratropical transition phase (i.e. 6th to 8th October) will lead to a windier solution over the UK. Difference tracking (Figure 2.6b) is saying that the run with a deeper storm over the UK located the depression further east (more progressed) at the earlier time. Note that ESA is suggesting that perhaps the more northerly element to the track of the depression may not be important.

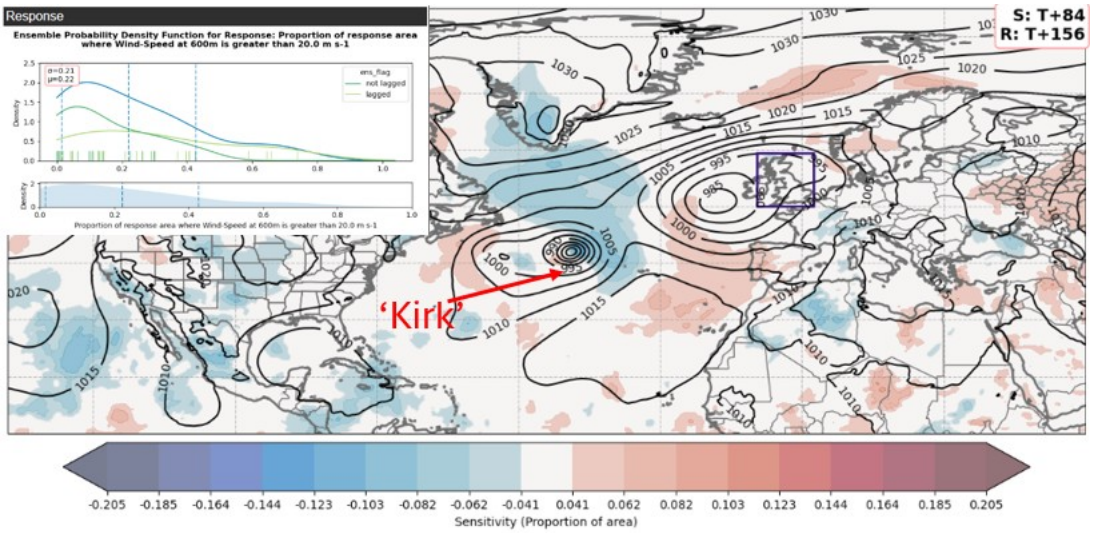
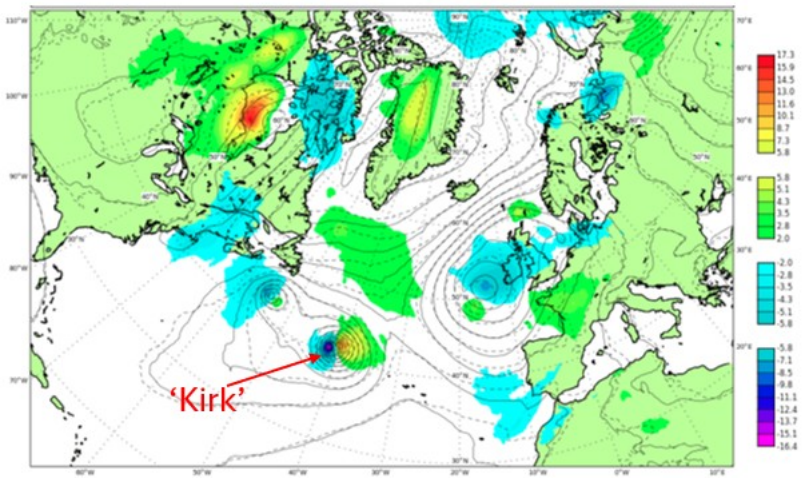


Figure 2.6a: ESA output from 00Z 4th October run. The insert at top left of this figure shows the Response PDF and provides meteorological context, highlighting that the earlier (lagged) run offered a windier solution. For this example, we have selected windspeed >20m/s over the UK valid at 12Z on 10th October (T+156) as the response function and MSLP as the sensitivity variable. The sensitivity time shown in the sensitivity plot is 12Z 7th (T+84). The black lines show MOGREPS-G mean MSLP, Colour = sensitivity, in the units of the response.



© Crown copyright 2024, Met Office

Figure 2.6b: Difference in MSLP between the GM DT00Z GM 4th October and the previous DT12Z 3rd October run valid 12Z 7th October. Solid black lines MSLP DT00Z 4th Oct, dashed lines MSLP DT12Z 3rd Oct, colour difference (DT12Z – DT00Z).

Although some of the insights provided by ESA (Fig 2.6a) are similar to those inferred by difference tracking (Figure 2.6b) ESA is directly linked to the location, time and response function of interest, i.e. it implicitly links to the chosen response functions or 'weather' that have impact. Difference tracking shows how much the forecasts differ at each grid point, which might or might not be relevant for later UK weather, while the ESA tells us where differences matter in relation to the weather of interest such as strong winds over the UK at a later time. While difference tracking is complimentary and offers important strengths, such as setting the scene and helping focus attention on areas seen in the ESA output, it can show differences that do not directly influence the weather variable of interest, for instance it shows differences in the depression to the west of Ireland, but ESA doesn't see that as being relevant for the subsequent windiness over the UK.

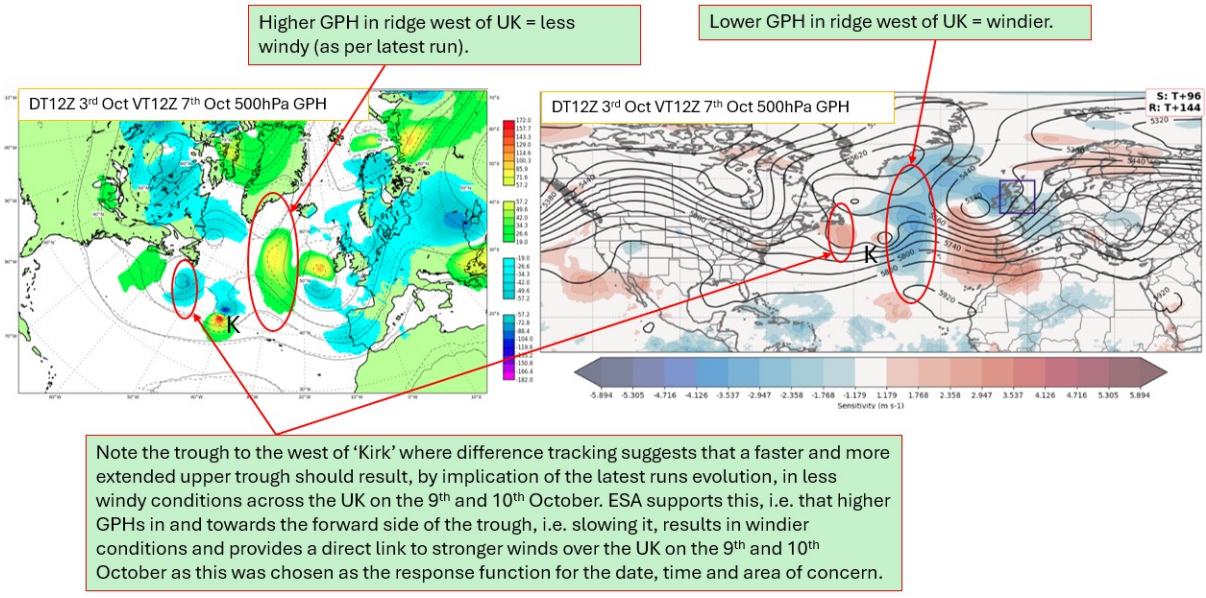


Figure 2.7: Model runs showing 500hPa Geopotential Height (GPH) valid 12Z 7th October 2024. 'Kirk' (K) lies between the two areas of interest highlighted above. Difference tracking on the left is showing the difference between the12Z (black lines) and the 00Z (dashed lines) runs of the GM on 3rd October. Blue shades = negative difference, green/red positive difference. On the right the ESA tool, for the response function windspeed >20m/s, is telling us in which direction GPHs need to change at a location to lead to a windier solution. Blue shades that lower GPHs result in a windier solution, while red shades indicate that higher GPHs lead to a windier solution.

Moving to 500hPa the ESA shown in Figure 2.7 (right) is telling us that at 12Z 7th October (48 to 72 hours before the time of interest) lower 500hPa Geopotential Heights (GPHs) (blue shading) east of 'Kirk' leads to windier conditions over the UK later 9th October and into the 10th but that the ridge forward of it, the trough to the rear of it and the position of the major extratropical vortex west of the UK

© Crown copyright 2024, Met Office

are also critical to the evolution. This suggests that the exact position of ‘Kirk’ might be less important than the other environmental features that will ultimately interact with ‘Kirk’ to provide a windy outcome over the UK many hours later. This is a key point in ESA, it highlights that sometimes the system itself at earlier times isn't as relevant as the controls around it.

The ESA shown in Figure 2.7 makes sense, as changes to 500hPa GPH in the directions indicated will result in ‘Kirk’ moving east and then being steered northeast as a deepening feature to pass near to, or across, the south of the UK. In a scenario where the 500hPa ridge/ trough couplet, lying approximately 35N to 45N, is more amplified, ‘Kirk’s’ eastward progress is resisted allowing the zonal trough extension of the vortex south of Iceland to cyclonically rotate around the vortex and extend south forward of ‘Kirk’ and hence steer ‘Kirk’ southwards. This can be seen in the runs of the 4th October where the northerly flow is stronger, the surface depression over the UK is maintained and ‘Kirk’ steered east and then eastnortheast (rather than northeast) on a more southerly track and failing to deepen.

The ESA tool can then be used to track these sensitivities back in time through the flow to identify their origins. This can provide greater insight and help to focus user attention on the factors in the early stages that matter to the outcome of interest. In this case ESA provides support to the interpretation formulated by looking at the differences between the GM runs shown in Figure 2.6b and Figure 2.7 (left). However, it also significantly increases our understanding of the evolution and what the drivers of the uncertainty are (and notably filters out those that aren’t important). For instance, utilising the ESA animation capability reveals the importance of the extratropical vortex west of the UK and the upper trough west of ‘Kirk’ around the 7th and 8th October. Notably it is also directly linked to stronger wind speeds by the user choice of response function (the parameter we are interested in).

Investigating the upper air evolution in the ESA output in Figure 2.8 also highlights the 300hPa GPH jet stream south of Greenland as a major player. Recall that the extension of this jet into Iberia was noted earlier as a possible source of uncertainty in the forecast for late 9th October and on 10th October. Here ESA confirms this and tells us that a south shifted jet around T+96 leads to a west shifted and veered jet (compared to earlier runs) that is associated with ‘Kirk’ taking a more southerly and less developmental track.

© Crown copyright 2024, Met Office

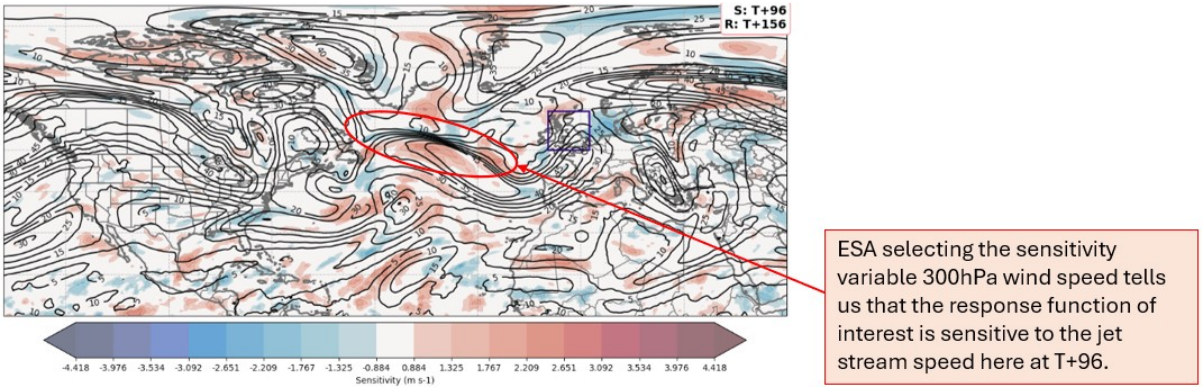


Figure 2.8a. ESA of the sensitivity variable wind speed at 300hPa, valid 00Z 7th October (T+96), for the response function wind speed >20m/s over the UK at 12Z 9th October 2024 (T+156).

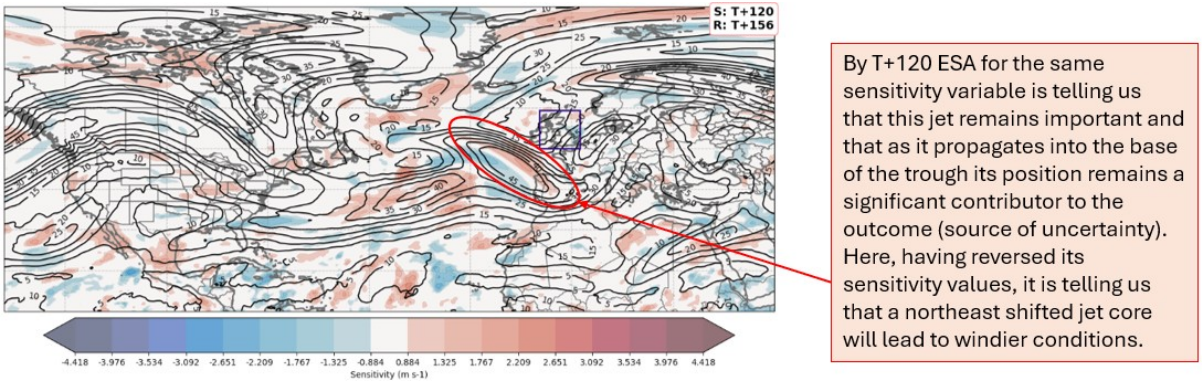


Figure 2.8b: ESA of the sensitivity variable wind speed at 300hPa, valid 00Z 8th October (T+120), for the response function wind speed >20m/s over the UK at 12Z 9th October 2024 (T+156).

Tracking the jet core highlighted in Figure 2.8, that has been identified as having a direct influence on the windiness over the UK late on the 9th October and into the 10th October, back to the forecast analysis time is shown in Figure 2.9. This demonstrates how ESA can be used to animate backward and forward in time to track features back to the early forecast evolution. In this case the animation tells us that, at T=0, the trough that enters the Atlantic at T+72 originates in the area highlighted by the red oval. The evolution of this trough is also influenced by the trough to its west over the Pacific. In later runs the jet rounding the base of the Pacific trough, which is also more extended, is stronger and enters the wave train over North America, splitting into a northerly (the jet highlighted in figure 2.8) and southerly stream accelerating the progression of the downstream ridge and the eventual arrival of the trough into the Atlantic forward of it.

© Crown copyright 2024, Met Office

ESA can be animated backwards and forwards to enable tracking of features with time. Tracking the trough that influences the evolution of 'Kirk' (see Fig 2.7) back through the sequence shows us that it originates over the US/ Canada border at T+0. More conventional difference tracking can be used to provide support and also shows (as ESA prototype domain is constrained to enable timely output) that a trough over the western Pacific feeds into the downstream evolution.

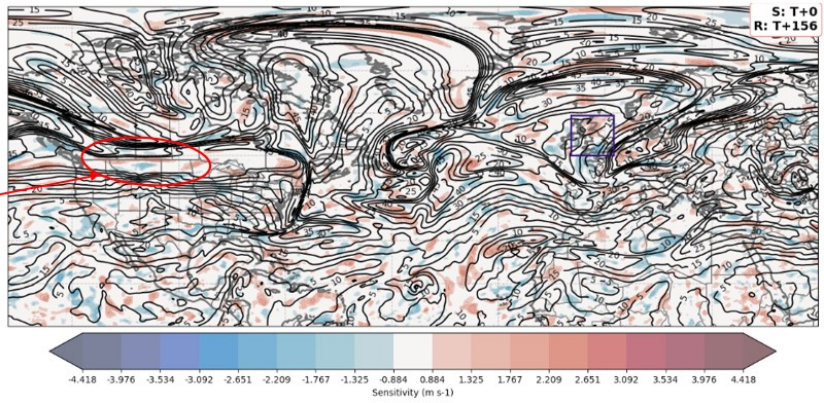


Figure 2.9: ESA stepped back to track the trough, identified at T+72, over the Atlantic to the west of 'Kirk' in Fig 2.7 to its origin at T+0.

The options for addition and selection of response functions and sensitivity variables are near limitless, and we have used only a few of the options selected for trial in the prototype tool to investigate this case. However, they have provided significant insight into the evolution and increased our understanding of what is important to the outcome of interest. In this example ESA draws one's attention to several features that are influential to the evolution of Hurricane Kirk as it undergoes extra-tropical transition and merges with the mid-latitude flow. This enables the user to begin to make informed choices by monitoring observational trends and evaluating quality of fit of deterministic and ensemble membership to early forecast stages. In turn this will enable earlier guidance and advice upon which decisions can be made to be provided.

Throughout this case study we have compared with difference tracking results, as a complementary diagnostic, and see similar areas highlighted (although the sign of the values for difference tracking are reversed due to the more recent run being the reference). Difference tracking adds support to our conclusions here and is an important part of the overall information package and forecast process employed to aid decision making. The strength of ESA is that it is spatially, temporarily and parameter specific allowing us to identify features of interest early in the forecast integration that have an influence over the parameter and response function that has impact. From this information, and inspection of analysis increments, observations and model trends etc. available on 4th October we can conclude that the recent model runs are indeed the more likely outcome and the guidance and advice from the EWH can be adjusted to move with greater confidence towards these solutions and at an earlier time.

© Crown copyright 2024, Met Office

2.1.5. Key findings from the ESA trial

Project team insights

The ESA project team have met fortnightly throughout 2024 and discussed and documented ESA cases of interest. Our main insights are summarised in this section. An earlier project report by Deployable Scientist Daniel Etheridge (Etheridge, 2024) provides a more detailed assessment of several of the key findings discussed below.

The ESA tools have been shown to be viable and insightful, giving a way to understand the “dynamical story” and highlight the relevant features driving the uncertainty in the ensemble for a given weather event of interest. ESA has shown it can provide useful information on precursors and disruptors in the model evolution that can provide insight for providing advice and guidance and, if an outcome remains uncertain, when more confidence will arise in the forecast solution. The 300hPa wind response function allows the dynamics to be interpreted in terms of jet orientation and/or position, which enhances understanding beyond the geopotential height view.

An interesting aspect of ESA is to study the “sensitivity of sensitivity” - investigating how the sensitivity field changes and depends on adjustments we make to input parameters. Early in the project we ran for two response areas in ESA on the day (a larger and smaller area centred on the UK) to investigate the sensitivity to the response box size. We found that sensitivity varied significantly for some cases where the larger box meant the response values were coming from weather systems in a different part of the UK from the area of interest and which had a different causal mechanism resulting in significant differences in the sensitivity patterns. The smaller UK area was preferred and taken forward, and the ability to set off bespoke runs via the On Demand interface for even smaller areas was created, to allow the sensitivity calculations to be homed in on the specific weather of interest.

A run-to-run analysis of sensitivity for a particular event revealed that some areas remain stable through time, and some change, reflecting changes in the response function data as the forecast evolves. Some degree of run-to-run differences are to be expected. They are just a reflection of the ensemble and the changes data assimilation provides, which itself changes the dynamics. Looking at the change in sensitivity with decreasing lead time for many high-impact cases would be a good next step to providing insights into the usefulness of sensitivity at different lead times operationally.

We have seen in many cases that sensitivity areas are rather widespread, even when close to the response time. This is due to dynamical coupling of the whole evolution of a forecast and can thus be considered indirect, rather than direct,

© Crown copyright 2024, Met Office

sensitivity. In other words, direct sensitivity highlights the features that are dynamically influencing the response, while indirect sensitivity shows areas that are related to direct sensitivity through ensemble covariance relationships. While both direct and indirect sensitivity can be valuable in an operational forecasting environment, the direct sensitivities are the most dynamically relevant and generally produce higher correlations with the response. When getting further away from the response geographically, it becomes harder to characterise what is direct sensitivity, indirect sensitivity and what can be considered noise, which is produced through spurious correlations due to a small sample size. This is one of the reasons why the ESA tools are specifically being developed with the EWH in mind, where the expert Op Mets have the skills to know which of the flagged sensitive areas are likely to be dynamically linked.

A useful tip for interacting with ESA is to start the sensitivity animation at the time of the response, and then track the sensitive areas back through the forecast. The key sensitive areas often track several days back in a forecast. The sensitivity fields tend to start becoming less distinct at around 24-hour lead time and lower, which is a potentially limiting factor for ESA-based on-demand sub-setting, and it would be interesting to examine whether this would be different if MOGREPS-UK were used in ESA.

Interpretation of ESA can sometimes be tricky, with cases where people may interpret the output differently. This highlights the need for ongoing training/support to develop/spread best practice, as should also be the case for all tools. Using ESA alongside other investigative tools and methodologies aids interpretation of the sensitivity output and should be encouraged. If several tools/methods are consistent in their findings of the key synoptic precursors to a particular event, more confidence can be attributed to that information.

Feedback from the Expert Weather Hub

The ESA on the day and ESA on demand tools have been trialled by the Expert Weather Hub (EWH) from July to October 2024, with feedback collected via three main methods: i) the August EWH Model Office; ii) a questionnaire via a Microsoft form; iii) in person feedback. The main themes emerging from each are given below.

<u>i) Model office</u>

The Model Office was an initiative run in the EWH in August 2024, where additional Op Mets were working in parallel to the on-shift Op Mets, to allow them to engage in experimental ways of improving forecasting practices. During each Model Office afternoon, a group activity was performed to explore the prototype ESA tools. A summary of the feedback received is provided below:

© Crown copyright 2024, Met Office

- Many participants asked if a quick reference guide in how to interpret the outputs could be produced, particularly in the early days of use or for future users.
- Participants felt it would be useful if examples could be provided of how the tool has been useful, as well as any cases that may have led to poorer guidance.
- Many participants highlighted a desire to select a previous submission to pre-populate the ESA on demand form so that it can be re-run with slightly different parameters whilst others remain consistent. For example, retaining the same area, response function and sensitivity variable, but using a different model cycle.
- It would be useful for the run date and time to include the day of the week i.e. Thu, or to use a pop-up calendar to select the day.
- Adding other ensemble systems was requested by several participants who felt the 36-members of MOGREPS-G may not sufficiently cover event space.
- Some participants suggested additional on-demand response variables to include CAPE and CIN, and 850hPa WBPT as part of the on-the-day sensitivity variables.
- On some events, there was a clear ‘event horizon’ identified where a dipole emerged at a future timestep and then persisted until the response time. This showed promise in understanding when a particular event may become more predictable, which is an encouraging result.

ii) Questionnaire

A Microsoft Form questionnaire was created and sent to the Op Mets in the EWH, asking them about their use of the EE prototype tools. Eight responses were received, and all had used the ESA on demand and on the day tools. All respondents had found the tools to be either somewhat useful, useful or very useful (i.e. no-one had said it wasn’t useful), and most had found the interfaces easy to use. Results were more split with regards to how easy it was to interpret the output, with three Op Mets finding this challenging. Responses to a series of more open questions have helped drive the recommendations in section 2.3, and are summarised below:

What did you like about the ESA tools?

- Several respondents praised ESA on demand, and the ability it gives to run custom response functions in areas and at times they define, allowing for a targeting of periods and features of interest, and for testing of various hypothesis.
- ESA had generally helped make it easier to identify the source of uncertainties and sensitive areas, especially after a little practise at interpreting the output. For some respondents it mainly confirmed pre-existing thoughts, but for others it added new insights, and gave information on when predictability of an event might improve.
- The webpages were praised as being straightforward to use, and quick and responsive.
- Several people commented that they had only scratched the surface of its potential

© Crown copyright 2024, Met Office

and so were keen for it to continue so they could try it more often.

How could the ESA tools be improved?

- Including multi-model ensembles to better sample the range of possible solutions was requested by several respondents, together with expanding to a global domain, and having the output available sooner after the model run time to enable its use as a time critical decision-making tool.
- Some people requested that ESA on demand by adapted to allow multiple ESA combinations to be submitted at once, and for the previous inputs to be remembered to allow later runs with tweaks.
- An interesting idea was whether a novel way could be developed to use the sensitivity data to trigger certain thresholds for an alert that there was an unusually sensitive region in this ensemble cycle.
- A couple of respondents discussed the potential of using ESA to identify a more accurate ensemble subset / narrow down which ensemble members tell the best story.
- Further training and training resources were requested, to help avoid misinterpretation and to enable new users to learn how to use it quickly and efficiently in an operational environment:
    - Examples of cases where it had highlighted unexpected links, or provided information that would have added value to the forecast process.
    - Cheat sheets and video demonstrations slowly explaining the ESA process and giving examples of common use cases.

What additional value did the ESA tools bring to your forecasting processes?

- The Op Mets praised that it allowed a quick analysis of the ensemble forecasts over large regions, allowing for a rapid and more reliable assessment of a huge volume of data.
- ESA had provided greater synoptic awareness of the key drivers of the forecast, enabling better understanding of predictability barriers and when they might be passed.
- One respondent replied that ESA had allowed them to identify sensitivity at earlier lead times than they thought were possible using just difference tracking across deterministic solutions.
- Another said that after using ESA they had a better understanding of which ensemble members were more reliable in the time frame that they were interested in.
- Two users had found that ESA mainly confirmed ideas they had already formed based on analysis on the broadscale pattern and anticipated meteorological processes.

Could you share any example cases where ESA has helped or supported your understanding of the ensemble evolution and synoptic drivers for an event?

- Three examples were given and are quoted below:
    - *“I once was trying to understand the return of a warm plume for a thundery breakdown. I found that early in the day the greatest sensitivity of the front returning was how far S the front pushes 2 days beforehand! But later in the same day (12 hours later), the sensitivity appeared to become more tied to amplitude of the upper pattern in the Atlantic, and the influence of the fronts clearance two days earlier lessened by that time.”*
    - *“The track of ex-hurricane Kirk was a case in point, where the trend to a more southerly track was dependent on various shortwaves. The ESA tool highlighted this well, and was very useful in lending confidence to the less impactful track.”*
    - *“Ex-Ernesto system - a more complex example where either an upper trough or surface low northwest of the hurricane proved quite critical in how the extra-tropical transition of Ernesto took place and what impact there was downstream on UK weather. Perhaps this would not have been as obvious in traditional NWP visualisations.”*

© Crown copyright 2024, Met Office

iii) Detailed feedback from an active user

We have also received detailed feedback from Nick Silkstone, one of the Expert Op Mets in the EWH who has regularly used the ESA on demand tool, and has helped other Op Mets by putting together a useful “ESA cheat sheet”:

“Meteorology is often very specific, and the response function and area selected is often critical to the sensitivities that emerge. Therefore, I found the ESA on the day tool which used a standard box over the UK as a response function to be too broad for my requirements. Therefore, I quickly migrated to ESA on demand, and have used that ever since.

It's been interesting learning how to define response functions, and areas, with a little trial and error. But I now have a good idea where to start for future events, for example using 850hPa WBPT as response function for tracking sensitivity to well-marked frontal zones etc. Many of the sensitivities revealed have been expected, but I’ve often been surprised to see how early they emerge as small sensitivity areas (I’ve found several cases where the sensitivity emerges 24-48 hours earlier than I thought through difference tracking), and then these can be traced to higher sensitivity regions at shorter lead times. I’ve also seen one event where the sensitivity appeared to run through various troughs and ridges with the group velocity at about 25° longitude per day all the way from North America to the UK in a strongly zonal flow, that was really cool to see. Of course, there are lots of unrelated sensitives that show up on these plots, but as an operational meteorologist you understand the bounds at which phase and groups velocity can travel, such as how we assess water vapour imagery. But I’d always want to see all the sensitivities plotted and used my trained eye to eliminate the one of no importance.

By chance I’ve not yet had the chance to use the tool around any medium / high impact events, but from what I’ve seen so far, I believe it is a very useful tool in the correct hands. In addition to helping identify sensitive areas, it has given me ideas around thresholds of predictability and at what point these may be crossed and predictability become much improved for some event occurring in the next week or so. Now it is something I’ve got, I would be very sad to lose this really insightful tooling.”

2.2. Ensemble Sub-setting (ESS)

2.2.1. Introduction to ESA-based Ensemble Sub-setting (ESS)

Ensemble forecasts try to cover the spread of possible outcomes; each member should be an equally likely representation of a scenario that might occur. However, the ensemble takes time to run and there is time between runs where additional information such as a new analysis or observations will become available. Using this additional information, we should be able to see which ensemble members are drifting from what has actually happened. The ensemble

© Crown copyright 2024, Met Office

can then be updated to exclude those members with the greatest error, or weight the members to refine the probabilities so that more weight is given to those members which more closely matched what has already happened, with the assumption that this will give us a better forecast at a later forecast time. It is unlikely that any member will be the best match at every point across the globe, so the sensitivity fields from ESA can be used to identify the points that are significant to the specific aspect of the future forecast that is most of interest. The early forecast errors in each ensemble member can then be calculated at those points where the sensitivity is greater than a chosen threshold, and ESA-based sub-setting carried out by re-weighting or sub-setting based on these focussed forecast errors.

2.2.2. ESS methodology implemented in EE WP1

To examine the viability of the ESS method we have, for simplicity, used the T+0 control member of the next MOGREPS-G run as our new analysis data. In reality, we would more likely use a high-resolution analysis or observational data; With this in mind we have used the satellite water vapour data to match against the pseudo MOGREPS-G water vapour forecasts where this is the sensitivity variable used within ESA. The ensemble sub-setting currently runs 4 times a day, with the timings reflecting when new MOGREPS-G analysis data becomes available for use (note if this were to be implemented operationally analysis/imagery data would be available much earlier):

- 00UTC analysis c.09 UTC, doing ESA-based sub-setting from previous day's 12UTC run
- 06UTC analysis c.15 UTC, doing ESA-based sub-setting from the same day's 00UTC run
- 12UTC analysis c.21 UTC, doing ESA-based sub-setting from the same day's 00UTC run
- 18UTC analysis c.03 UTC, doing ESA-based sub-setting from the same day's 12UTC run

ESA-based sub-setting is carried out for each of the response functions that form part of ESA "on the day", as was described in Table 2.1. The results for the responses using PCA are excluded from the display and objective verification as these would need to make use of the EOFs which are not routinely stored. The steps in the ESA-based sub-setting for each response function are as follows:

i) The sensitivity fields at the analysis time which were calculated in ESA "on the day" for that response are used to select sensitive areas at which to calculate the forecast errors in each ensemble member to inform the ensemble sub-setting. These points are selected to be those where the absolute sensitivity value is greater than a threshold e.g. 50% of the max

© Crown copyright 2024, Met Office

absolute sensitivity value across the data. The standardized sensitivity field has been used as this is what is in the ESA display but it might be more mathematically justifiable to use raw and this should be looked at in any future work.

ii) At these points, for each ensemble member, we calculate the differences in each sensitivity variable between the observation/analysis and the forecast from the previous model run at that validity time. Two different calculation methods are being considered to summarise the ensemble performance across these points:
    1. MAE (mean absolute error) is the mean of the absolute difference values across all the selected points.
    2. PROJ (the projection method, as defined in Ancell (2016)) is designed to take full advantage of the sensitivity fields, using the sensitivity values to weight errors i.e. errors in areas of larger sensitivity should have more influence. At each selected point, the error is multiplied by the sensitivity (this value could be negative or positive and represents the estimated change to the response from the error at that point based on the sensitivity). These products are then summed across all selected points to yield a holistic estimated change in the response. By retaining the signs of the error multiplied by the sensitivity, competing dynamical effects are accounted for in the estimated change in the response.

iii) Each of these methods produces a value for each ensemble member which can be ranked from smallest to largest, with smallest being the best fit and largest the worst fit. Other difference algorithms such as FSS could also be investigated in the future.

iv) This ranking can then be used to subset or reweight the ensemble to hopefully give an improved ensemble forecast for the response in question. It is worth noting that as this subset is tailored to a specific response, it would be inappropriate to use it more widely, and subsets for different responses are likely to be different.

ESA-based sub-setting has been running from August to October 2024, producing output to illustrate the sub-setting approach to forecasters (described in section 2.2.3), and outputting data to allow objective verification of the ensemble sub-setting (described in section 2.2.4).

2.2.3. Description of the ESS prototype display including illustrative case

Output from the ESA sub-setting is meant to be illustrative of how the sub-setting works but not how the sub-setting would eventually be used. For the ESA sub-setting trial display, we have selected the following options:

- a threshold of 50% of the maximum absolute sensitivity value to identify the points

© Crown copyright 2024, Met Office

- the PROJ method for the calculation to summarise the ensemble performance across the sensitive points
- the whole domain from the sensitivity plots as the domain to be considered in the calculation
- the 18 and 9 best-performing ensemble members are selected as sub-sets and plotted in the prototype display, in the form of PDF and meteorological context plots.

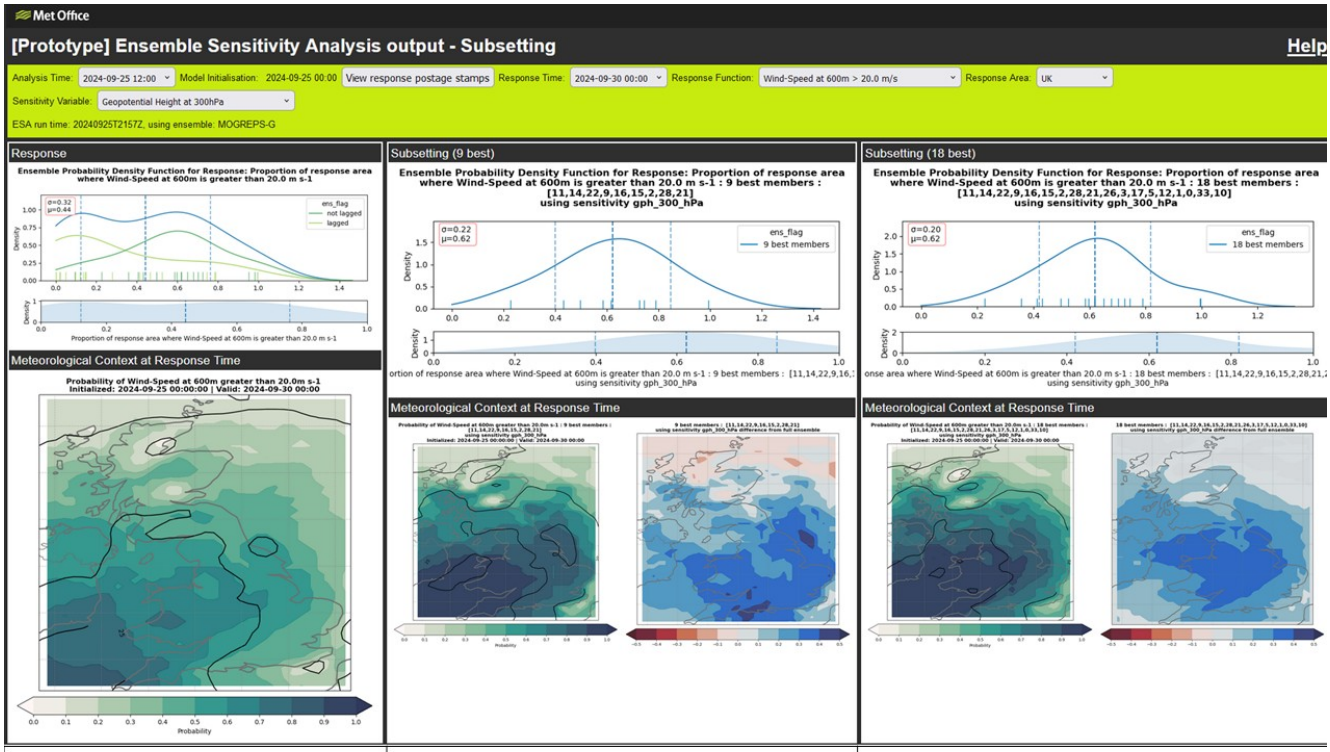
Ideally in a future implementation of ESA-based sub-setting, there would be more flexibility on the area the sub-setting is calculated over, how to re-weight or select subset size, and how to display the output. The current display was created to give an initial illustration of the sub-setting concept to the Op Mets in the EWH in near real time.

Figure 2.10 shows a screen shot of the ESS user interface. At the top of the page are drop down menus to select different output, described in Table 2.2. Underneath the drop-down menus are three panels that display the original and sub-setting output for the selected options. Underneath the screenshot of the user interface in Figure 2.10 are descriptions of what is contained in each panel.

Analysis Time	This is the time of the analysis/observation data used in the sub-setting.
Model Initialisation	This is the model run used for the ESA outputs and ensemble forecasts. This is usually the model run 6 or 12 hours before the Analysis Time. The value is automatically updated when the Analysis Time is changed.
Response times	These are the same as those available in ESA on the day.
Response parameters	These are the same response parameters that are used in ESA on the day, with the exception that the PCA response parameter is not currently available.
Response Area	A box covering the UK – the same as is used in ESA on the day.
Sensitivity variables	These are the same sensitivity variables as in ESA on the day. For each sensitivity variable the ensemble rankings, and hence the selected subset, will be slightly different. The different subsets for each sensitivity variable can be viewed by selecting the sensitivity variable from the drop-down menu.

Table 2.2. Explaining the drop-down menus in the Ensemble Sub-setting output page.

© Crown copyright 2024, Met Office



Left-hand panel shows the original response function (the PDF plot with the values for each of the 36 ensemble members as green dashes). This is identical to the response function in the ESA on-the-day display. Below it is the original Meteorological Context image.	**Middle panel** shows the response function re-drawn using the 9 best ensemble members calculated by the ESA sub-setting method. Below it is the revised Meteorological Context image using just the 9 members and the difference between the original Meteorological Context image and the image using only 9 members (9 members minus the original ensemble).	**Right-hand panel** shows the response function re-drawn using the 18 best ensemble members calculated by the ESA sub-setting method. Below it is the revised Meteorological Context image using 18 members and the difference between the original Meteorological Context image and the image using only 18 members (18 members - original ensemble).

Figure 2.10. Screenshot from the Ensemble Sub-setting (ESS) user interface. In this case both the 9- and 18-best member subsets have shifted the PDF to the right, with a higher proportion of the response area having wind speed greater than 20m/s in both subsets (as shown by the blue colours on the difference plots).

2.2.4. Objective evaluation

The ensemble sub-setting method has the potential to narrow down the possible outcomes for impactful events as new information becomes available. To establish whether the method works we carried out objective verification to answer the following questions:

- Does ESS give an improvement, on average, over the full ensemble forecast?
- Does using the ESA in the sub-setting give improvement over just picking the members with the smallest errors or the latest model run?
- Which of the many potential configurations would give the best results:
  - Which sub-setting calculation (MAE or PROJ) gives the best results?

© Crown copyright 2024, Met Office

- What sensitivity threshold should be used for including points within the sub-setting calculation?
- How does the choice of sensitivity parameter affect the results?
- How does the number of members used in the sub-setting affect the results?

The amount of ESS data we have at the end of this project is limited and as with all ensemble data, large volumes are required to accurately verify the skill of an ensemble. In the section below we will discuss the methodology we used to objectively verify the usefulness of the ESS method and any limited conclusions we can make from the data available. This will hopefully provide a basis for more extensive research into ESS once more data is available, assuming a decision is made to consider using this method further.

Evaluation method

For each of the responses and sensitivity parameters within ESA on the day (OTD), the sub-setting processing produces the member rankings and corresponding error values using the MAE and PROJ methods. This processing is being done using two cutoff thresholds, 50% and 30% of the absolute sensitivity values within the domain, to test the impact of varying this threshold. The domain used is the area in the ESA OTD plots. Ideally, we would limit the domain to areas where the sensitivity is most influential on the response rather than including indirect sensitivities and possible noise. However, to do this accurately would require intervention by an Op Met to pick that area. In addition to the 50% and 30%, we are also testing a 0% threshold i.e. all points within the domain are included. This allows us to measure how much additional benefit is gained from using ensemble sensitivity to define the points to include in the sub-setting calculation, rather than just taking the ensemble members with the minimal error across all grid points in the domain.

To evaluate the performance of the subsets we need an observation, which is taken by calculating the response value using T+0 MOGREPS-G control member, to give a measure of the truth for that response function. This is defined as the observation as this is the response value that was observed at the response time.

Using the ensemble rankings any subset could be created. Moving forwards, it would be possible and useful to investigate varying subset sizes and/or re-weighting based on the errors. However, to keep the initial evaluation relatively straightforward, we have created several subsets of the response values containing 9 and 18 ensemble members, to correspond with the data shown in the display. The subsets we are testing are as follows:

- The original ensemble
- MAE 18 best members

© Crown copyright 2024, Met Office

- MAE 9 best members
- MAE 9 worst members
- PROJ 9 best members
- PROJ 9 worst members
- A random set of 9 ensemble members
- First 9 members (this uses the ensemble members 0 to 8 making it a challenging subset to beat as it contains the control and other members from the latest run)
- MAE 18 worst members
- PROJ 18 best members
- PROJ 18 worst members
- A random set of 18 ensemble members
- First 18 members (this uses the ensemble members 0 to 17 (the control and other members from the latest run))

These different ensembles of response values are compared to the observed response value. Many metrics could be used to compare the relative skill, but here we start with two simple metrics, for both of which the smaller the value the better:

i) Mean square error of the ensemble mean compared to the observation (MSE).
ii) Continuous Ranked Probability Score (CRPS). This is a standard ensemble verification metric which measures the difference between the ensemble CDF and the observation CDF. It is equivalent to the Brier Score across all thresholds. This metric should reward smaller spread (sharpness) if the forecast is accurate, which is important when considering ensemble subsets.

Timeline plots are also created showing the error of the ensemble mean from the original ensemble and the subsets to illustrate the performance of sub-setting on a day-to-day basis.

Initial results of the ESS evaluation

An example timeline plot is shown in Figure 2.11, in this case for the T+48 ensemble mean error of the 95th percentile wind speed at 600m, from the original ensemble and the subsets formed from the 18 best and 18 worst members using the MAE sub-setting method and a sensitivity variable of 300hPa geopotential height. The performance varies significantly from day to day, with some indication of cases such as 00UTC on 23/08/2024 where the best member subset gives a significant improvement, but some cases where it doesn’t lead to an improvement. Overall, in this example, the subset with the 18 best members performed better than this original ensemble in 60% of the cases, and this rises to 67% when only the higher wind speed cases (> 20m/s, shown in grey shading) are included.

© Crown copyright 2024, Met Office

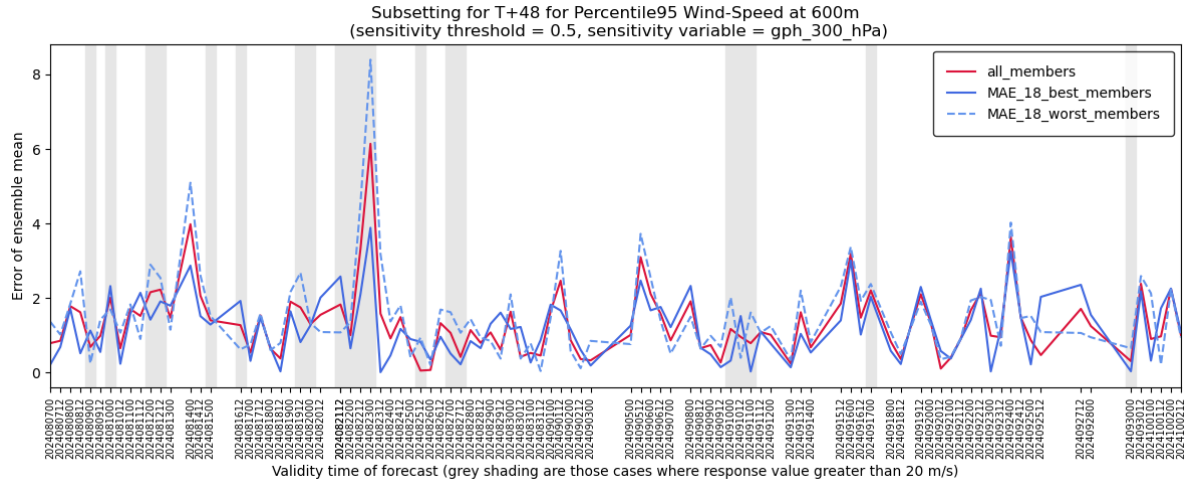
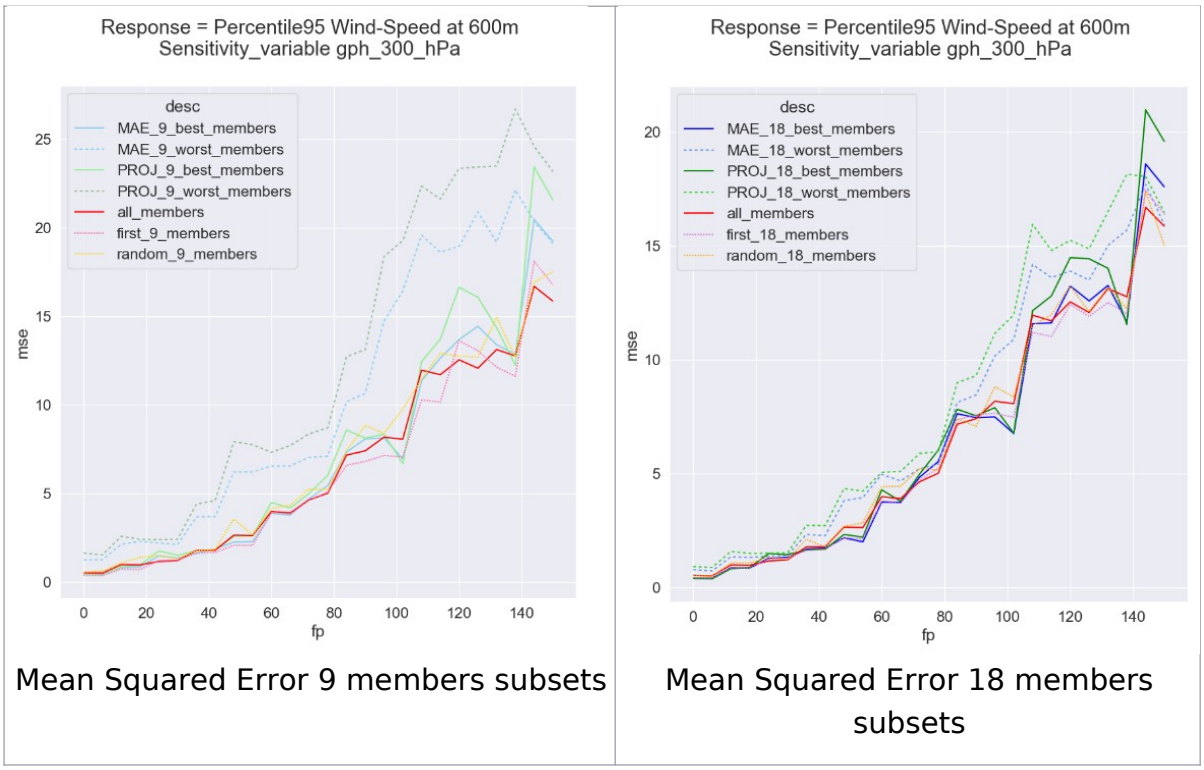


Figure 2.11. Timeline of the errors of the T+48 ensemble mean from the original ensemble and the subsets formed from the best and worst 18 members, for the following ESA combination: Response function = 95th percentile wind speed at 600m, Sensitivity Variable: 300hPa Geopotential height.

Identifying case by case examples using these timeline plots is interesting, but in order to more thoroughly assess the performance of the sub-setting, the remaining objective verification is carried out across all cases. To make things simple we will again look at just one set of options:

- Response is 600m Wind-speed 95th percentile
- Sensitivity parameter is 300hPa geopotential height
- Sensitivity threshold is 50% of max absolute sensitivity within the domain



Mean Squared Error 9 members subsets

Mean Squared Error 18 members subsets

Figure 2.12. Mean Squared Error against forecast period for the different subsets described in the evaluation method.

© Crown copyright 2024, Met Office

The example in Figure 2.12 shows the Mean Square Error (MSE) of the ensemble mean compared to the observed (mse) against all of the forecast periods (fp). As would be expected, the error increases with forecast period for all the subsets and the original ensemble. The graphs are not very smooth, due to the small sample of data currently available. The main things that are clear is that the sets of worst members seem to be performing worse than the other subsets. There appears to be more error with the 9 members compared to the 18 members. The sets of best members, random members and first n members are all very close together especially at the early lead times. There is currently not sufficient data to draw any definite conclusions on these more subtle differences. It is a similar story across the other responses, sensitivity parameters and sensitivity thresholds although there are some subtle differences in the performance across different responses and sensitivity parameters. Similar plots are shown for the CRPS in Figure 2.13 and tell a similar story.

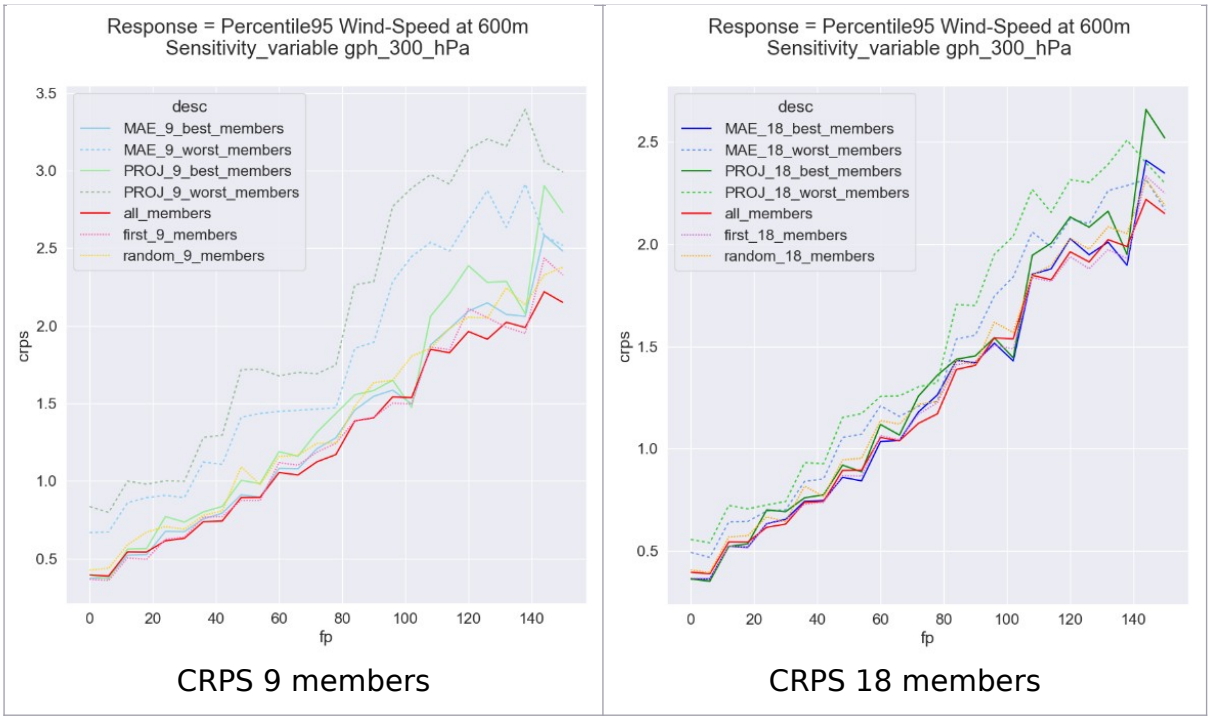


Figure 2.13. CRPS against forecast period for the different subsets described in the evaluation method.

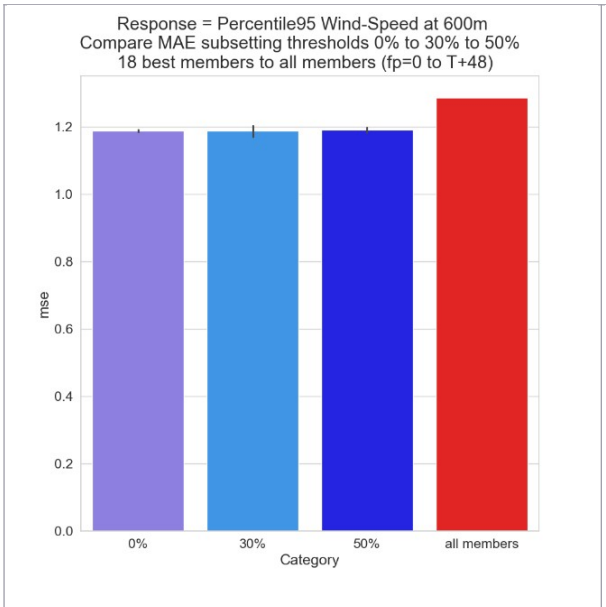
To try and smooth the data, the forecast periods are combined into three time periods: i) T+0 to T+48; ii) T+48 to T+96; and iii) T+96 to T+144. In this report only the plots from the first period are included, as this is where the most benefit is expected to be found, but the plots for the other time frames are available from the ESA subsetting evaluation webpages. The next few sections use this combined data to examine the performance of different aspects of the sub-setting.

How does changing the sensitivity threshold affect the performance of the sub-setting?

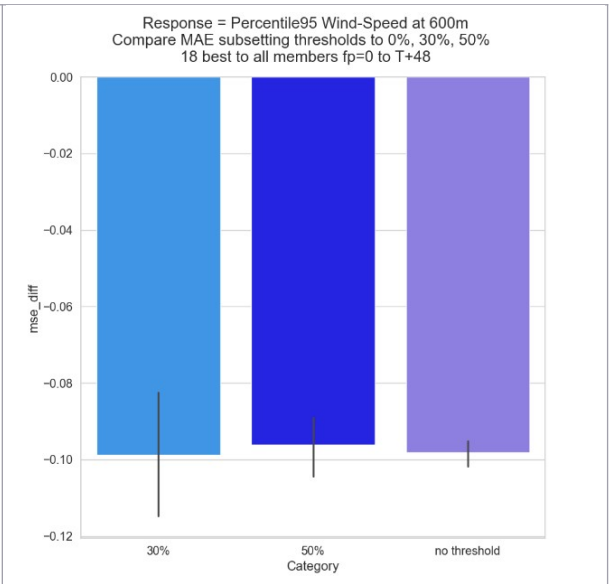
© Crown copyright 2024, Met Office

The box plots in Figure 2.14 display the MSE averaged across all cases, and all lead times T+0 to T+48, for all members and several MAE 18 member subsets calculated using different sensitivity thresholds. The MSE is slightly lower (better) than the full ensemble in all the subsets. If we look at how changing the sensitivity threshold affects the MSE, the left-hand chart in Figure 2.14 shows that there is very little difference. To make it easier to compare in the right-hand chart in Figure 2.14 we have calculated the difference between the MSE scores of the different thresholds to the value for all members of the original ensemble. In this plot negative values are better than the original ensemble, positive values are worse. This makes it easier to see the differences, but care needs to be taken as the relative difference are still very small. We will use this difference between the performance of the aspect of the sub-setting and the performance of all ensemble members in Figure 2.15 as well.

The error bars on the plots in Figure 2.14 are across the different sensitivity parameters (excluding water vapour, which had similar results but wasn't available for all runs). As we can see this does not make a great difference in the overall values but could make the difference between which version performs better than the other. However, there isn't sufficient data to make any definitive choices as to the best performing sensitivity threshold to use in the sub-setting. We will limit the plots in Figure 2.15 to just using 300hPa geopotential height as the sensitivity parameter and the 50% sensitivity threshold.



Varying the sensitivity thresholds for MAE method, 18 best members and comparing to all members (verification statistic is Mean Square Error (mse), where lower is better)



Again, varying the sensitivity thresholds as in the previous plot but now calculating the difference between the results for each threshold and all ensemble members (mse_diff -ve values better than all members, +ve worst)

© Crown copyright 2024, Met Office

Figure 2.14. Effect of varying the Sensitivity thresholds.

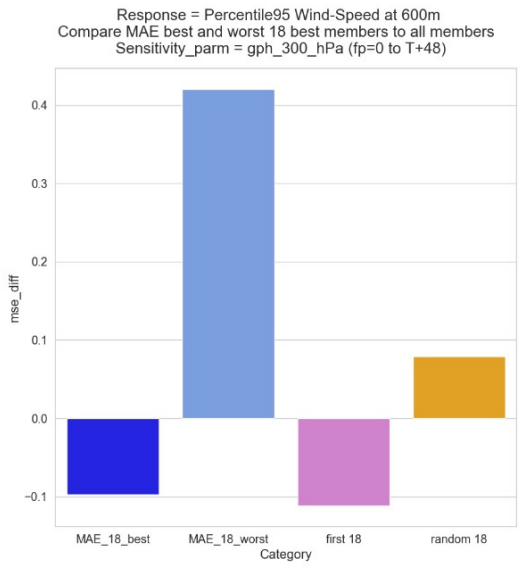
Examining performance of ESA-based sub-setting against the full ensemble and other subsets, and comparing using 9 vs 18 members and MAE vs PROJ method.

The top row of Figure 2.15 examines the performance of ESA-based sub-setting against the full ensemble and other subsets when using both the MAE method (top left) and the PROJ method (top right). The results reflect what was shown in Figure 2.12. The best 18 members do better than the worst 18 members for both the MAE and PROJ method (this is the same across almost all responses and time periods). The best members usually outperform the random members, and the original ensemble. The first 18 members (i.e. the 18 members from the latest model run including the control) is a hard subset to beat, and does slightly better than the 18 best member subset in this example, but this isn't always the case. The differences show that neither the MAE or the PROJ method only pick the latest 18 members. The bottom left plot shows that the best 18 members usually outperform the best 9 members. The MAE method seems usually to be better than the PROJ method but this could be due to the use of standard sensitivity rather than raw sensitivity. These results indicate that excluding the worst n members when creating a subset might be useful.

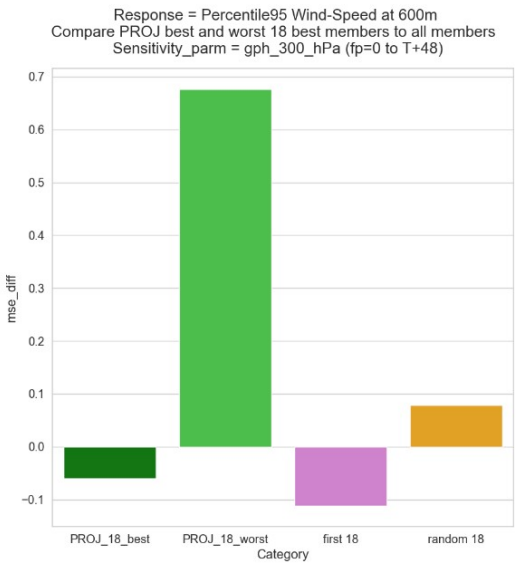
Difference in Mean Squared Error between "all members" and subset (mse_diff)

-ve values are better than "all members", +ve values are worse than "all members"

MAE 18 best members v worst members. Included are the results for the first 18 members and a random set of 18 members.



PROJ 18 best members v worst members. Included are the results for the first 18 members and a random set of 18 members.



© Crown copyright 2024, Met Office

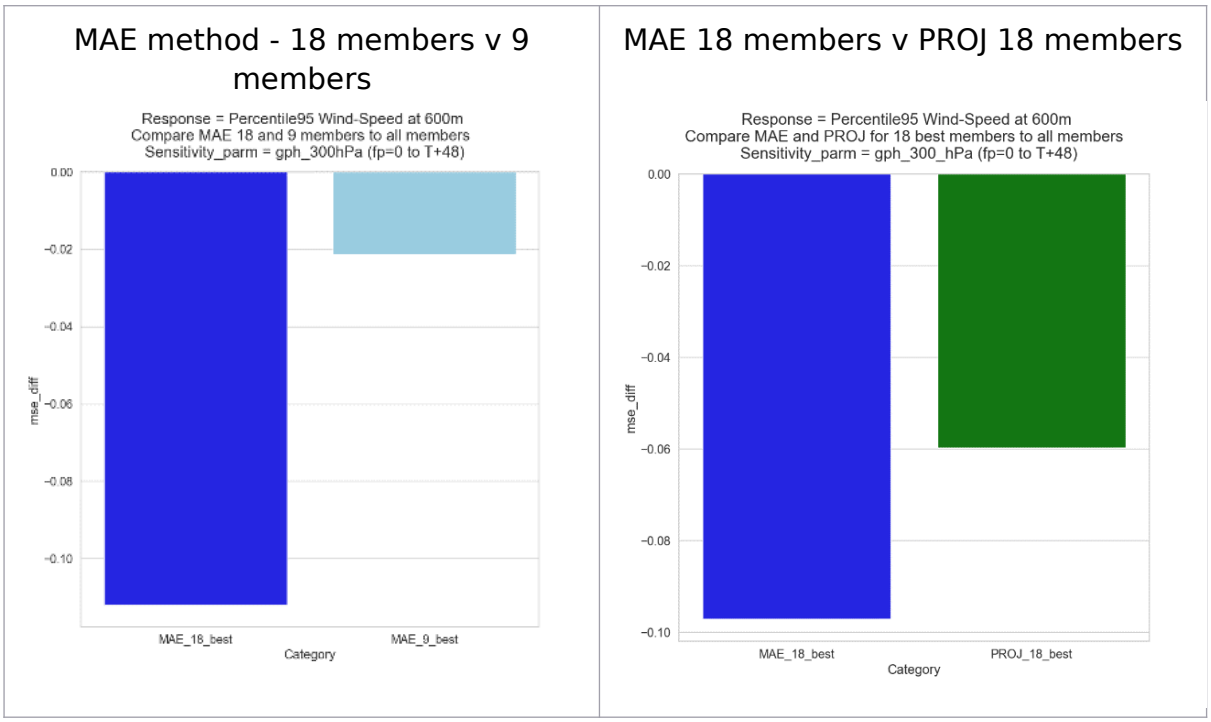


Figure 2.15. Looking across the different aspects of the subsets.

In summary:

- We have carried out an initial objective evaluation of ESA-based sub-setting, using the full ESA sensitivity domain to identify the sensitive areas used in the sub-setting.
- The system does appear to have some skill in identifying the worst members as they score quite a lot worse. However, at the moment, the differences between the full ensemble and the remaining sub-sets are too small to be sure that this method would lead to improved forecasts.
- In order to get robust results, and be able to conclude whether ESA-based sub-setting is adding significant value, this analysis needs to be repeated when more data is available. Further investigation of the optimal sub-setting setup including Op Met selection of areas, is also recommended.

2.3. Recommendations for ESA and ESS

A set of recommendations for the future development of ESA and ESS have been collated and agreed by the ESA project team, which range from short to long term and from small incremental changes to more complex ideas for future developments.

Short-term recommendations for the prototypes:

- Keep the ESA/ESS prototypes running through until March 2025 for the EWH to continue to trial during the winter months.
- Continue the work to scope operational development of the ESA tools for use

© Crown copyright 2024, Met Office

by EWH, as ESA has been shown to provide valuable insight. An initial task would be to re-evaluate which responses variables, functions, and sensitivity variables are required.
- Scope the ESA training required to support best practice including building a repository of examples/best practice to enable Op Mets to confidently use and interpret ESA.
- Allocate ongoing EWH resource for running ESA cases in real time in support of the on-duty Chief/Deputy Chief. The user guide should also be updated to include updated tips on how to interpret sensitivity output.
- Use ESA to help to build a dataset of ensemble characteristics (strengths, weaknesses, situation-dependent confidence etc) by giving a greater understanding of why the ensemble is behaving as it is in a given scenario.

Ideas for enhanced technical capability:

- It may be possible to only require an operational ESA on demand, rather than needing both ESA tools, if ESA on demand is enhanced to allow batch (several response and sensitivity combinations run together for the same area) and repeat (additional runs submitted later for identical area) running.
- Add capability to define a response over an area during a given time frame e.g. 95th percentile of wind speed within a 24-hour period.
- Develop a multi-model ensemble version of ESA including EC ENS and NCEP GEFS. Due to the underspread nature of MOGREPS-G, there will be times where the true solution falls outside of the spread, and in these cases ESA will be of limited use, as it can only work with the solutions contained within the ensemble. Adding the multi-model capability will increase the spread in solutions and improve the value of ESA.
- Expand the tools to be available on a full global domain.
- It would be useful if the ESA display could allow for more than one run to be displayed and to be scrolled through side by side, perhaps for two response functions or for the same response function but for differing iterations of MOGREPS-G.
- Develop “ESS on demand”, allowing the ESA-based sub-setting to be performed on a smaller area of choice, rather than the full domain, in a similar way to ESA on demand:
  - Any lessons learned from the Image Comparison prototype that was developed under Data Triage to carry out ensemble re-weighting based on satellite imagery will be important to consider here, alongside other comparison methods.
  - Ideally there would be flexibility for Op Mets to tailor the subset, and plot probability difference fields for a response function between the full ensemble and a derived subset. ESA-based sub-setting is tuned to the specific response function that it was derived from, and it could be helpful to aid understanding if the Op Met could see if the subset has resulted in

© Crown copyright 2024, Met Office

large changes in another variable.

Further forecasting development ideas:

- Examine if/how ESS could bring value and be used in reality in EWH working practices.
- Investigate how the ESA/ESS and the feature-based clustering could be linked e.g. looking whether the sub-sets are focussed on a particular cluster or are spread across them, which might indicate whether some clusters look more likely to occur than others.
- Apply representative member selection technique to sub-sets to help deliver storylines.

Further scientific development ideas:

- ESA:
  - Investigate sensitivity post-processing, including localisation, smoothing, or filtering.
  - Analyse the climatology of sensitivity, looking at how it varies with location, time of year, and meteorological conditions.
  - Investigate the potential to apply ESA to other areas of science, or in the model development and evaluation process. A technical separation of the core ESA code from the OTD/OD processing could help make this easier to do.
- Sub-setting:
  - Change to use raw sensitivity instead of standardized sensitivity in the sub-setting.
  - Repeat the ESS objective evaluation once more data is available to increase the robustness of the results. Rank histograms could be a useful addition to evaluate whether more events are falling outside of the ensemble spread in the sub-sets.
  - Evaluate if improved sub-setting performance can be achieved from i) varying the subset size based on the error values i.e. cutting the ensemble when the errors reach a certain level; ii) reweighting each ensemble members using their corresponding error values to redistribute the PDF to favour the members with the least error.
  - Objective sub-setting experiments could be performed to investigate how to optimise sub-setting configurations in different meteorological setups, and see if there are certain situations e.g., large ensemble spread, specific response distributions (such as Gaussian or bimodal) within which sub-setting is more effective.
  - If ESS on demand can be set up, verify the outcome from these Op Met led sub-set selections, to help reach an overall conclusion as to whether ESA-based sub-setting has potential to add value to the forecasting process.

© Crown copyright 2024, Met Office

3. Difference tracking

Difference Tracking is a technique that calculates differences between two model runs with the aim of disclosing the departures in their forecast evolution. By evaluating how the differences evolve throughout the forecast period the goal is to identify how the more instrumental synoptic features affect the development of the forecasts.

In this section we compare and contrast the results of the difference tracking techniques with the ESA so as to address their compatibility in an operational environment. We then go on to discuss the results of a feasibility study which has been carried out to apply the difference tracking techniques within the MOGREPS-G ensemble.

3.1. Comparison Of Difference Tracking and Ensemble Sensitivity Analysis

3.1.1. Introduction

An important aspect of the work of the Expert Weather Hub (EWH) is to provide guidance on the dynamical drivers of the current forecast period and to give insight into the atmospheric evolution and on the reliability of the forecast. A fundamental part of this process over many years has involved the use of Difference Tracking (DT) techniques which utilises the evolution of differences between two model solutions through the forecast to identify those dynamical evolutions which are critical to the forecast trajectory.

As part of the work undertaken in Work Package 1 of the Ensemble Exploitation Strategic Action, Ensemble Sensitivity Analysis (ESA) has been implemented at the Met Office for trial use by Op Mets in the EWH (see section 2.1). ESA is a technique which discloses the dynamical response in the forecast to changes in the initial conditions across all ensemble members by utilising the ensemble (covariance) statistics. It discloses the sensitivity of some chosen quantity (the ‘response’) to some other variable (the ‘sensitivity’).

Whilst the two techniques described above are clearly very different, there is also considerable synergy between them as they share the common goal of highlighting the areas of a model trajectory which are impacted through changes in the initial conditions and model physics. Thus, one could expect that there would be similarities between differences across members of the ensemble and differences between deterministic forecasts ran from similar initialisation times. As it is likely that both techniques will be utilised by the EWH for the foreseeable future, it is of importance that an intercomparison of the two is performed to

© Crown copyright 2024, Met Office

assess their relative merits and compatibility.

This section summarises the work undertaken to compare the output and guidance from the DT and ESA techniques. It presents the key findings of the intercomparison and illustrates their compatibility with a number of annotated examples. It shows that the two techniques are compatible with each other, and together provide a complementary means to the Op Met of gaining insight into the driving features of the forecast.

3.1.2. Comparisons of DT and ESA

The systematic study carried out here considered a number of forecasts over a period of several months. DT was used to disclose the areas of major departure between successive model solutions. This was then compared with the evolution of the sensitive areas disclosed by the ESA.

Figure 3.1 shows an example from the 00 UTC 16th September 2024 forecasts in which an anticyclone was forecast to build over the UK but then give way along its southern flank to the approach of a low-pressure system from the south towards the end of the forecast period. The approach from the SE of the low was associated with a forecast of thundery and potentially heavy rain and some degree of uncertainty with variable locations and intensities of precipitation predicted across the southern half of the UK. To illustrate the evolution of the sensitivities and differences, the ESA and difference tracking fields are shown at T+48 and T+84, with the difference tracking being the differences between the QG00 16th September 2024 and the QG12 15th September 2024 runs. Selected similarities between the two techniques that are assessed as being important to the boundary of the high and low pressure systems across the south of the UK are indicated by arrows at the two forecast times, and there is clearly a very consistent indication of differences and sensitive areas between the two diagnostics. In particular, the emerging sensitivities within the flow from the North Atlantic upper tough is clear in the ESA and this is replicated as significant differences in the difference tracking at T+48. There is also a good indication in both fields of sensitivities and differences across southern Europe associated with the large upper vortex there. Both diagnostics at T+48 effectively also highlight the merging of influences from the North Atlantic and Mediterranean as indicated by differences and sensitivities along the southern flank of the High. Both diagnostics therefore support each other and show that model uncertainties in the approach of the large thundery low to the SE are heavily influenced by dual developments both in the North Atlantic and across the Mediterranean. This increases confidence in the evaluation of the synoptic development for the Op Met.

An additional contribution to the development in this case is the influence from a

© Crown copyright 2024, Met Office

short-wave feature propagating towards the UK from the east (circled). This is indicated in both diagnostics but requires careful scrutiny of the fields and a knowledge of the context of the dynamical meteorology to identify its role. This is an example of how the two diagnostics support one another by disclosing features to varying degrees when other sensitive areas are disclosed within the vicinity whose role is unclear.

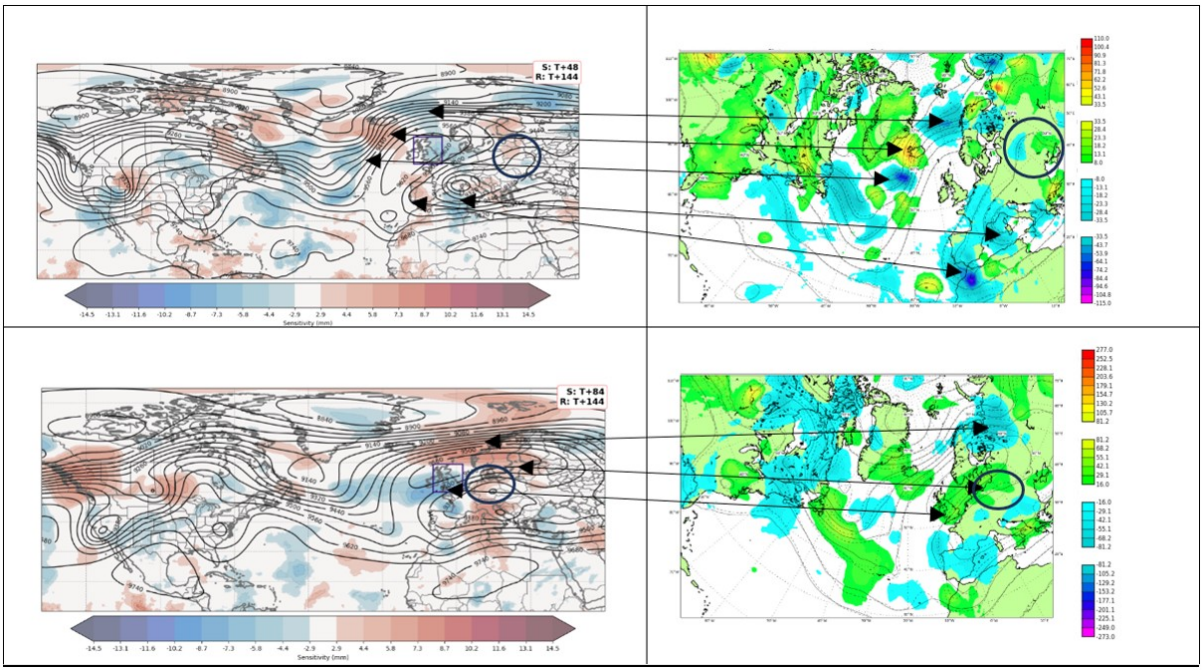


Figure 3.1. ESA (left) and difference tracking (right) example fields showing similarity of selected areas identified (arrows). The relative positive and negative signs of the comparison is not thought to be significant and merely reflects the differences in the techniques. Here for forecasts initialised at 00 UTC 16th September 2024. Top row: T+48, lower row: T+84. Response time in the ESA is T+144. ESA Response function: maximum precipitation in 24hrs. Sensitivity Variable: 300hPa gph. Difference Tracking is for 250hPa gph between QG00 16th September (solid line) and QG12 15th September 2024 (dashed line), colours: differences. For clarity only some areas have been shown which are pertinent for the development of the high and low interface over southern England, but many more similarities exist. More subtle but important sensitivities are circled and are supported in both diagnostics.

As another illustration of the comparison, the 00 UTC 17th July 2024 T+60 Global Model forecast indicated a short-lived heatwave over much of the UK with a large amplitude upper trough to the west and southerly winds advecting an increasingly high WBPT airmass over the country. A comparison of the ESA and DT diagnostics for this forecast is shown in Figure 3.2. The ESA at this time showed strong dependency on the ridge near the NW of the UK (with a stronger ridge associated with higher temperatures) and also on the developments across the North Atlantic (e.g. a deeper trough being associated with higher temperatures). DT for the same forecast with respect to the previous GM run showed a very similar indication of a dependency to the trough/ridge system both near the UK and out to the west, and perhaps more clearly emphasises the impact of the developments on the ridge across the south of the UK. However,

© Crown copyright 2024, Met Office

even with the overall similarities there are differences in the signals from the techniques. For example, the large amplitude upper trough to the west of Iberia is considerably more marked in the difference tracking compared to the signal in the ESA thus implying a more prominent role of this feature in the forecast evolution in the difference tracking. This case therefore further illustrates the overall similarity and compatibility of the two techniques, whilst emphasising differences in some development areas for the Op Met to consider with more scrutiny.

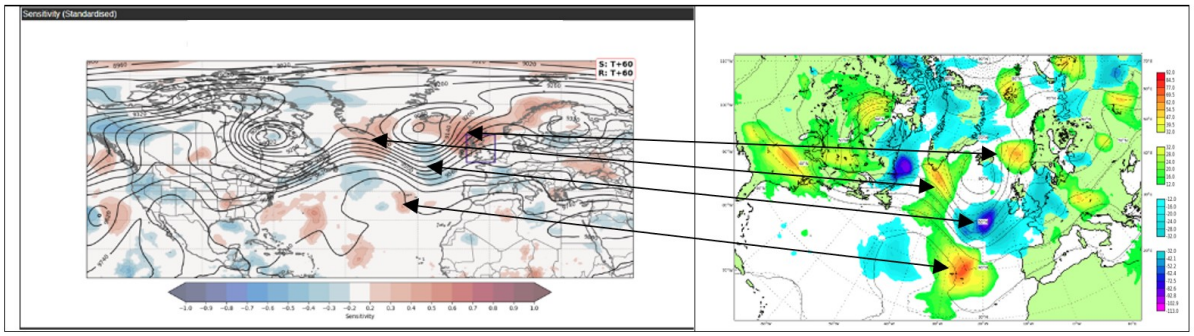


Figure 3.2. ESA (left) and Difference Tracking (right) for the QG00 17th July 2024 T+60 forecast. Some areas of Difference Tracking and ESA are highlighted but the similarity is very clear elsewhere also. However, despite the close similarities there are also some significant differences between the systems (e.g. west of Iberia) which can prompt further consideration by the OpMet. ESA Response is for mean screen level temperature at T+60, ESA Sensitivity Variable 300hPa gph. Difference tracking is for QG00 17th July 2024 T+60 (solid lines) – QG12 16th July 2024 (dashed lines). Colours: differences. Verification time for all fields is 12 UTC 19th July 2024.

3.1.3. Conclusions

The results of the work comparing the ESA and DT are summarised below:

1. Difference Tracking and ESA show considerable degrees of similarities, often highlighting very similar developments, especially when considered over broad scales.
2. These areas are not identical however, and there are often differences between the relative locations, extent, and magnitudes of differences and sensitivities. This is anticipated given the fundamental differences in the techniques. The differences between the two techniques provide a valuable way of highlighting different developments which perhaps can be overlooked using one method alone and bring additional useful insights.
3. All differences and sensitivities are internally consistent and dynamically relevant within each system. Thus the differences between them are considered a strength of using both methodologies together as they can highlight a feature in one technique which may be overlooked in the other, prompting a complementary re-evaluation across the two systems.

© Crown copyright 2024, Met Office

4. ESA is generally broader in appearance than difference tracking. There are a number of possible contributing factors to this including the statistical origins of the ESA, lower resolution of the ensemble (difficulty representing short-wave features), and the relative differences between members run from the same data time in the ensemble. Difference tracking, in comparison, is disclosing the reaction of individual deterministic model differences at higher resolution from different run times or different modelling centres. This means that difference tracking can at least appear to be more reactive than ESA to sudden changes in the model trajectory whilst the ESA generally appears more stable. Stability versus reactivity are neither strengths nor weakness of either approach but merely emphasise their complementary nature.
5. Due to the potentially more reactive nature of difference tracking it may be more capable of highlighting certain features which can be thought less important in the ESA in some forecasts. That said, ESA sensitivities can often be identified at much smaller magnitudes before some differences are large enough to identity confidently. This again underlines their complementary nature.
6. ESA sensitivities can be traced back in a similar way to difference tracking thus reaffirming the synergies between the two techniques.
7. The relative “sense” of the sensitivities (i.e. positive and negative) in the ESA and the differences in the difference tracking are not thought to be relevant. Whilst the techniques are both measures of model sensitivity they are entirely different processes and should be interpreted as such.

Within the context of operational forecasting, the suggested best practice would be to employ both methodologies with a view to assessing each technique’s areas of sensitivity/variability. An intercomparison of those developments identified in both techniques may highlight areas which were initially judged less important in one technique alone. This allows the Op Met to gain greater insight into the origins of model uncertainty and the drivers of the forecast.

3.2. Feasibility Study of Difference Tracking Applied to Ensembles

3.2.1. Introduction

Difference Tracking is a useful technique in Operational Meteorology to understand changes in model solutions between two deterministic model runs. It is usually applied between different runs of the same model from different analysis times or different runs from different modelling centres at the same analysis times. These differences are then used to aid in the identification of the largest differences between the two model solutions and to evaluate the

© Crown copyright 2024, Met Office

evolution of those differences in order to identify their origins.

Ensembles provide a set of forecasts (members) which are all valid from the same data time and have been perturbed from the analysis of the control to yield different solutions. These members have traditionally been processed to yield probabilities of some event occurring thus providing an indication of the sensitivity of the model to uncertainties in the model analysis and formulation. Yet the presenting of probabilities in this way takes away an opportunity from the Operational Meteorologist to evaluate the different member solutions to understand the underlying mechanisms involved in the evolution of the different members.

This feasibility study aims to address if difference tracking in ensemble space has any value in Operational Meteorology in the same way that difference tracking in deterministic models has been useful for many years.

3.2.2. The Feasibility Study

The first step undertaken was to write python code for the MOGREPS-G difference tracking as part of a new GitHub repository. The difference tracking was then applied to all the ensemble members with respect to the control to calculate the differences, i.e. member minus control. The differences are calculated on a small number of model variables that are deemed suitable for difference tracking from experience with the deterministic models.

The differences are presented in two ways; as postage stamps and as individual member minus control differences. The postage stamps view enables an overview of the relative size of the differences within the ensemble to be evaluated and also helps to identify any members with larger differences. The individual member difference plots then enable a more detailed examination to be carried out of the evolution of the member forecast that has been identified through the postage stamps.

Figure 3.3 shows an example set of forecast differences for the 00 UTC 3rd October 2024 run at T+72 for MSLP. This example has been chosen for illustrative purposes at a time when Hurricane Kirk is forecast to undergo extra-tropical transition during the forecast and in which the Global deterministic model indicates a trajectory to affect southern UK. At this time approximately one third of the non-lagged members indicate a similar solution of the storm affecting the UK (not shown), with the implication of possible strong winds and heavy rain. The difference tracking postage stamps reflect the spread in the ensemble members and highlight how various features have been perturbed from the control (shown in the top left of the figure). Whilst there are differences evident across the whole domain, there are a few areas which appear to occur more frequently: the location of the large depression near the UK (to right of

© Crown copyright 2024, Met Office

centre of the domain) and Hurricane Kirk in the mid-Atlantic (lower left of domain). This suggests that these features are more prone to uncertainty and have been more affected by the ensemble spread. The large UK low is characterised by a large degree of differences and shows both location and depth variation in the members. For Hurricane Kirk, the dipole of differences around the feature is also of varying positive and negative magnitude and location and orientation. This indicates that the speed and location of Kirk is highly uncertain with some members faster and some members slower with its progression towards the midlatitudes.

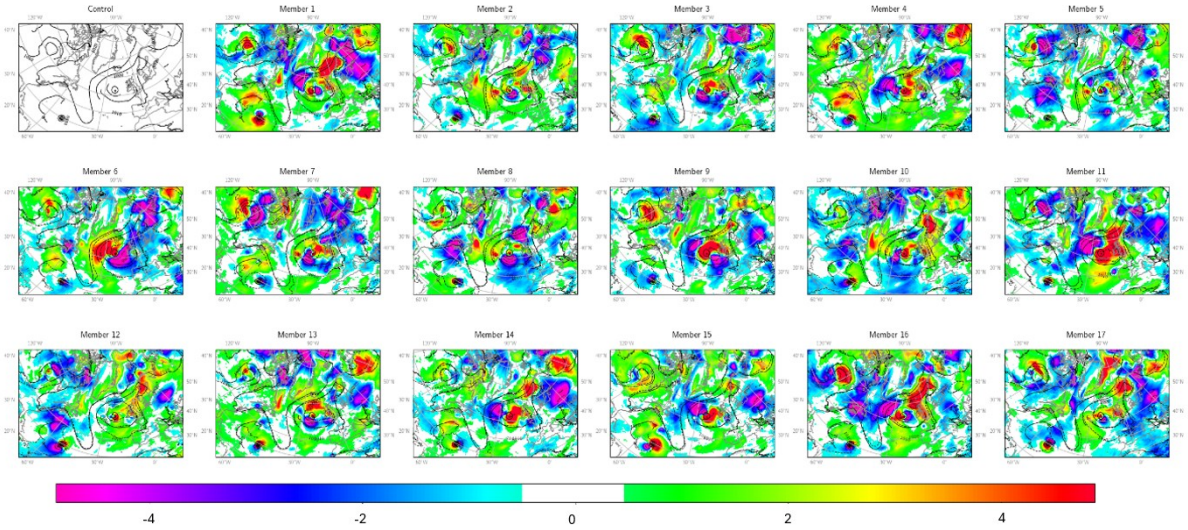


Figure 3.3. Example MSLP difference postage stamps for the 00 UTC 3rd October 2024 T+72 MOGREPS-G ensemble forecast. Solid contours: Control, Dashed Contours: Member, Colours: members-control difference. Valid time: 00 UTC 6th October 2024.

With the spread of differences disclosed by the postage stamps overview, this facilitates the evaluation of a particular member of interest in more detail. Figure 3.4 shows the differences between the control and member 1 identified above for 300hPa gph. 300hPa gph is shown here to more clearly disclose differences in the midlatitude flow. These differences highlight departures in the upper troughs upstream over North America, with differences in the amplitude and speed of progression evident. Differences in the shape and orientation of the vortex near the UK are also highlighted. Thus, the differences in the eventual evolution of Kirk in this member can be understood to a first approximation as originating from differences in phasing between Kirk and the upper troughs emerging into the North Atlantic, with an additional contribution from accuracy of the large vortex resident near the UK.

© Crown copyright 2024, Met Office

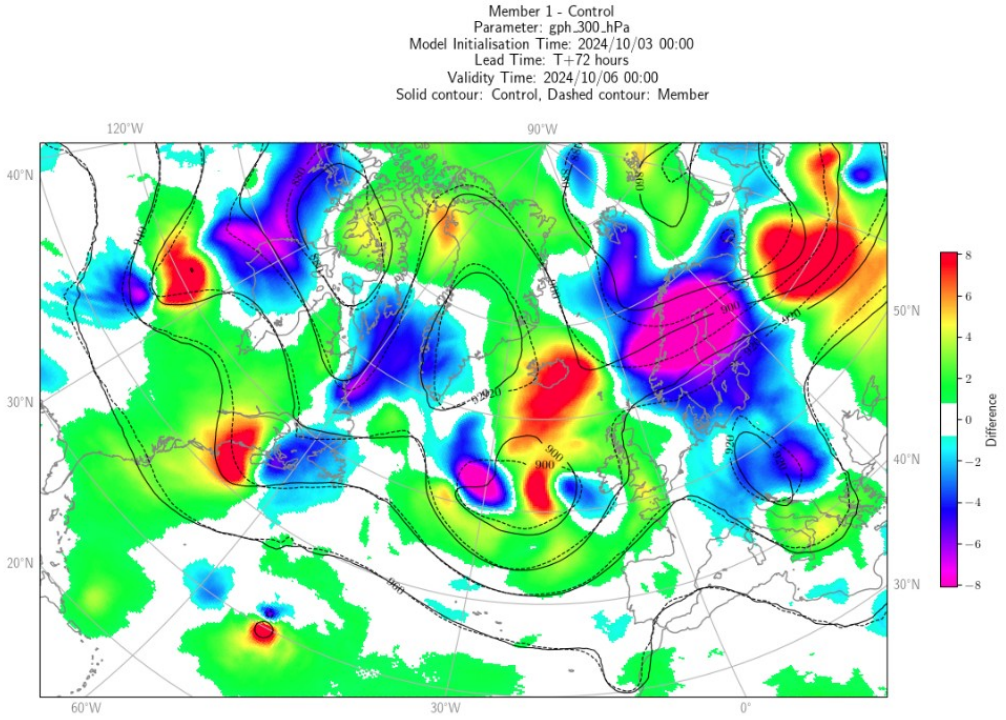


Figure 3.4. QG00 3rd October 2024 T+72 member 1 – control 300hPa gph differences at T+72. This highlights differences in the location of Hurricane Kirk in the lower left but also highlights differences in the amplitude and speed of progression of the upper troughs across North America.

3.2.3. Conclusions

This feasibility study has shown that difference tracking within MOGREPS-G is a viable way for gaining insight into (a) the differences in evolution that generate the spread, and (b) the differences in evolution that generate a particular solution. The pursuit of (a) is possibly more challenging in an operational environment if the spread is large (and unnecessary if the spread is small), but (b) would be achievable if a particular member of interest was identified. This member of interest could be identified by direct consideration of the postage stamps as discussed above or through other data sources such as the identification of a representative member using the clustering approach described in Section 4. In many cases there may be quite subtle differences in the ensemble member solution and difference tracking will be more difficult and it may be that other methods such as ESA are more suitable. However, in more extreme cases such as shown here the differences are easier to follow and track and are a viable diagnostic tool for Op Mets.

3.3. Recommendations

- It is suggested that difference tracking and ensemble sensitivity analysis are two complementary processes for Operational Meteorology and should be used as such. An intercomparison of the developments highlighted in both

© Crown copyright 2024, Met Office

techniques in the operational environment provides a broader means of evaluation of the synoptic evolution and allows a greater potential for gained insight. It is also suggested that the use of one technique is likely to help in the interpretation of the other.

- It is suggested that difference tracking in MOGREPS-G has shown potential value for Op Mets, and in particular fits with the production of guidance in the EWH where an understanding of the key drivers in a forecast is fundamental to the role.
- Whilst this feasibility study has shown potential, significantly more work is required to make it more suitable for operational use. The MOGREPS-G difference tracking should therefore remain and be maintained in R2O to be developed further as a research tool. This will over time provide a non-operational but accessible real-time facility to the Op Mets but also provide a useful research tool for case evaluations in research.
- There is potential to expand the range of ways in which difference tracking is applied to ensemble forecasts. For example, the output data from the feature-based clustering (as discussed in section 4 of this report), includes representative members from the clusters to help provide storylines when discussing likely forecast scenarios. The existing code developed in this feasibility study could be used to study the differences between the control and the representative members in interesting cases, or further developed to allow difference tracking between the various representative members.

© Crown copyright 2024, Met Office

4. Feature-based clustering

Feature-based clustering groups ensemble members according to their common evolutions for a particular weather feature over a window of interest, then identifies representative members from each cluster to give distinct forecast evolutions with associated probabilities, to help with communicating forecast scenarios to users.

In this section we give the background to this novel clustering approach, detail the methods involved, and show example outputs as have been trialled by the EWH in this project. We end with recommendations for how to take this work forwards in the future.

4.1. Background and introduction

Ensemble clustering is an efficient ensemble post-processing approach that can distil an ensemble forecast into its prevalent forecast scenarios by grouping similar ensemble members together – something which is increasingly important in a world where ensemble data volumes are rapidly increasing. The application of a suitable clustering method combined with appropriate forecast visualisation allows a forecaster to effectively focus on the key possible outcomes, simplify the message and characterise and communicate forecast uncertainty more easily.

Traditional ways of interrogating the ensemble information when assessing scenarios and risk of high impact events are through (1) probability maps and (2) percentile maps. Probability maps are useful if you have a customer who is only concerned about a particular parameter threshold being exceeded (e.g. 24-hour rainfall > 25mm), but they lose information from the full probability distribution and make it harder to assess interdependencies between multiple variables. Percentile maps are useful in estimating best- and worst-case scenarios (such as represented by the 5th and 95th percentiles for example), but the resultant gridded fields do not represent physical solutions from the model. This is where the new feature-based clustering described in this section becomes useful, as it will group members according to their common evolutions over a predefined window of interest, before identifying central members from each cluster which will hopefully represent distinct forecast evolutions which are suitable for communicating to the end user.

Several clustering methods already exist, which can be split into two main camps:

1. **Clustering of ensemble members according to a set of predefined climatological weather patterns or regimes**. Examples include the Met

© Crown copyright 2024, Met Office

Office's medium- to extended-range probabilistic weather pattern forecasting tool called Decider (Neal et al., 2016 and 2024), and Meteo France's short-range clustering product using a set of climatological rainfall and wind patterns (Mounier et al., 2022 and 2023).

2. **Clustering directly on the ensemble members.** In the literature, this typically involves k-means or k-medoids clustering of ensemble members according to their similarity to the leading modes of spatial variability within the ensemble using Empirical Orthogonal Function (EOF) analysis. Examples include the ECMWF regime-based clustering product (Ferranti and Corti, 2011 and Ferranti et al., 2015), and NOAA's ensemble clustering product (Lamberson et al., 2023).

Met Office Op Mets already take advantage of clustering output provided by Decider and the ECMWF product for issuing forecast guidance within the medium-to extended-range period. At these lead times, these products assist with identifying key transitions in the broad-scale setup, with often very high confidence of a particular weather pattern or regime occurring within the first few days. However, although there may be high confidence at the synoptic scale, there may still be uncertainties at the more local scale in terms of the magnitude and timing of different weather hazards, such as heavy rainfall, strong wind gusts, or extremes in temperature. This is where some of the clustering techniques which cluster directly on the ensemble can come in and is the focus for the clustering research and product development carried out as part of Work Package 1 of the Ensemble Exploitation strategic action.

As part of this strategic action, the Met Office have developed and trialled a prototype feature-based ensemble clustering approach, following many of the methods initially developed by Boykin (2022) in her PhD work at the University of Reading and in collaboration with the Met Office. The methods presented by Boykin (2022) were initially trialled in real time at the Met Office during a winter 2022 testbed with encouraging results (Porson et al., 2022), with the methods presented in this report providing a further development of this work. In this novel clustering approach, k-medoids clustering is applied to Fractions Skill Score (FSS) distances between members (Robert and Lean, 2008; Skok and Roberts, 2018, Gilleland et al., 2020), using identified features in each member. These features may represent areas of hazardous weather or mesoscale or synoptic features, such as areas of heavy rainfall, damaging wind speeds, or weather fronts.

This new approach includes several clustering and post-processing steps (described in detail in Section 4.2), which are designed to (1) determine spatial distances between objects and cluster on those distances, (2) identify an optimal number of clusters, (3) identify windows of interest, and (4) identify representative members within each cluster and time window, to provide

© Crown copyright 2024, Met Office

plausible forecast scenarios or evolutions with associated probabilities.

In total, 18 clustering configurations using MOGREPS-G are running in parallel, which will need to be refined in the future based on user-feedback. These configurations are split across four parameters (850 hPa WBPT, 10 m wind gusts, 6-hour total precipitation, and PMSL), with a mixture of percentile and absolute thresholds used for each when defining features for clustering. Having multiple configurations allows the user to pick the configuration that best suits the weather situation at the time.

The live prototype clustering display went live in early July 2024 and can be viewed from within the Met Office at https://www-nwp/~wits/clustering/ (accessed 16/10/2024). A product user guide is also available from this webpage.

The clustering code and clustering product generation code is contained in two separate GitHub repositories at https://github.com/MetOffice/clustering and https://github.com/MetOffice/clustering_products (both accessed 16/10/2024). Code is written in Python and uses configurations files as input to allow for maximum flexibility in running a range of setups. Anyone within the Met Office GitHub organisation will be able to access the repositories and external collaborators can be added with relevant permissions.

4.2. Methodology

Clustering is based on k-medoids (Brusco et al. 2019), which is an iterative clustering method, whereby *k* represents the number of clusters. Initially, *k*-members are chosen at random to represent the cluster medoids, and then members are assigned to each medoid based on the member-medoid pairing with the smallest distance. The cluster sum of distances are then calculated and saved for later use. One medoid is then changed in each following iteration and members are allowed to move clusters if a closer medoid is found. The number of iterations is capped by the number of members. Once all iterations are complete, the final member-cluster assignment is based on the group of clusters with the smallest sum of distances between members and their medoids, summed across all clusters.

K-medoids clustering is very similar to k-means clustering (Omran et al. 2007). The main difference is that with k-means, the sum of distances in each cluster is based on the distances between each member and the mean of the cluster. K-medoids was chosen over k-means because (1) the centre point (medoid) is restricted to being a member of the cluster instead of a mean of members (centroid) – this works better for fields like precipitation where means smooth out too much information and don’t provide physical solutions; (2) k-medoids is quicker to run as there are only so many member and medoid pairings, whereas

© Crown copyright 2024, Met Office

k-means keeps recalculating the centroids, so it can take a while to reach a minimum distance between all members in a cluster and their closest centroid (often subject to a user-defined cap on iterations); and (3) k-medoids is more likely than k-means to recognize an outlier as a singleton cluster if it is significantly different to other members – something which is particularly useful for identifying low likelihood, high impact scenarios.

The k-medoids clustering can use any distance measure. Here, we use Fractions Skill Score (FSS) distances between member pairs (Roberts and Lean, 2008; Skok and Roberts, 2018) – this is used as a form of data dimensionality reduction before the clustering is carried out. The FSS is a neighbourhood verification measure for assessing the difference between two fields. It was chosen as the distance measure within the clustering due to its ability to cope with discrete or noisy fields. For example, with precipitation it will account for subtle differences in things like frontal positioning, shower location, and timing, allowing for two members with these subtle differences to be mapped together within the clustering.

The FSS distances are calculated using identified features in each member. Here, features are typically defined as anywhere greater than a predefined threshold, such as >10 mm for 6-hour rainfall accumulations. Percentile thresholds can also be used, where they are calculated from a relevant climatology using all grid points, lead times and members from the ensemble being clustered. The thresholding has the effect of masking the gridded data and producing a binary field of ones and zeros (where ones represent grid points which exceed a given threshold), which are passed through to the FSS distance calculations.

FSS distances between each member pair are calculated for each time-step, before being averaged across all time-steps within multiple overlapping pre-defined windows of interest. The trial configurations use 24-hour windows of interest which have a 12-hour overlap between windows. 24 hours was chosen as this will typically capture the passage of a low-pressure system or frontal system across the UK. The clustering is then applied to these averaged distances over each window, allowing for members that have a similar evolution over time to be grouped together. This is a key difference from Boykin (2022), who carried out clustering on each individual time-step, before grouping clusters over multiple time-steps where ensemble membership remained relatively stable. However, early live trials of that approach showed that cluster membership was too unstable to be useful operationally.

Finally, representative members (RMs) are derived from the cluster medoids and should provide distinct and physically representative forecast evolutions for each cluster over the full window of interest.

The clustering steps described above are repeated for several values of k (number of clusters). This is because at this stage, we don’t yet know what the

© Crown copyright 2024, Met Office

“optimal” number of clusters should be to fully capture the uncertainty and distribution of members within the full ensemble. With the current setup, clustering output is produced for cluster sets (using k=2, 3 and 4), before a decision tree is applied (Figure 4.1) to identify the optimal number of clusters, which may vary by clustering configuration and window of interest. The aim behind this is to identify the largest cluster set which has RMs that are sufficiently different from each other.

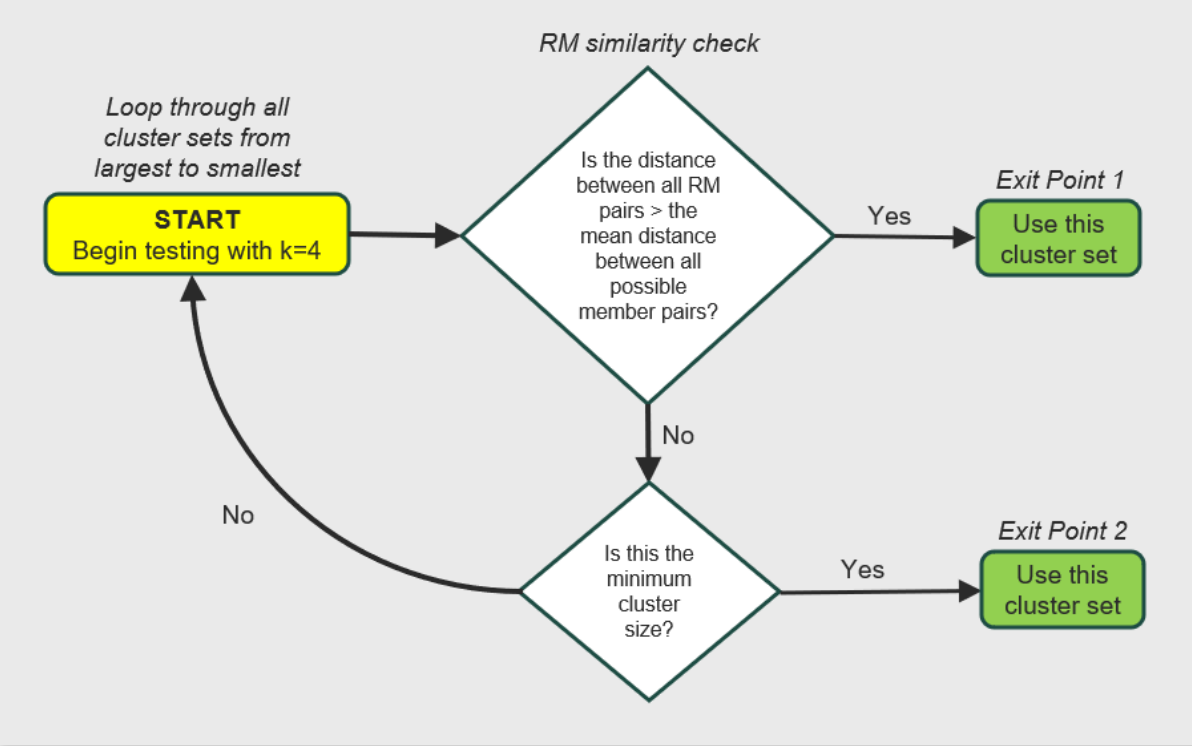


Figure 4.1. Workflow for identifying the optimal number of clusters for each window of interest.

Finally, product generation is carried out for the optimal cluster size in each window of interest. Forecasting products are initially presented for each window of interest, with more detailed forecast information available for individual time-steps within each window once a window of interest is selected within the top-level forecast webpage. This allows for a rapid assessment of all windows, as well as a "drill down" approach for the window of interest. The products produced are described in more detail in Section 4.3.

The multiple clustering steps described above and can be summarised in seven top level steps for a given configuration:

1. Identify relevant features in each member and time-step (binary).
2. Calculate Fractions Skill Score (FSS) distances between all possible member pairs for each time-step.
3. Calculate the averaged FSS distances for multiple overlapping pre-defined windows of interest.
4. Apply k-medoids clustering to the window-averaged FSS distances for several cluster sizes (values of k) – this is currently done for 2, 3 and 4

© Crown copyright 2024, Met Office

clusters.

5. Identify Representative Members (RMs) for each cluster size, based on cluster medoids (central members).
6. Identify the optimal number of clusters for each window of interest, from those produced in the steps above. This is based on the largest cluster set that has sufficiently different RMs (this is where the FSS distances between all RM pairs are greater than the mean distance between all possible member pairs).
7. Generate forecasting products.

© Crown copyright 2024, Met Office

4.3. Example clustering output

Example clustering output is shown for a rainfall case study on Friday 18th October 2024, using the time-lagged 36-member MOGREPS-G ensemble output from the 00 UTC run on Monday 14th October 2024. This example illustrates how the clustering can be used to distil an ensemble forecast down to a small number of distinct forecast scenarios over a predefined 24-hour window of interest.

The MOGREPS-G 24-hour postage stamps valid on Friday 18th October 2024 (Figure 4.2) reveal a lot of spread in solutions, with some members having heavy rainfall covering much of the UK (e.g. Member 5, Control (T-6), and Member 23 (T-6)), and other members restricting it to different parts of the UK. The members could be subjectively clustered by eye, but this is a time-consuming and subjective process, and so instead the objective feature-based clustering method described in Section 4.2 is applied.

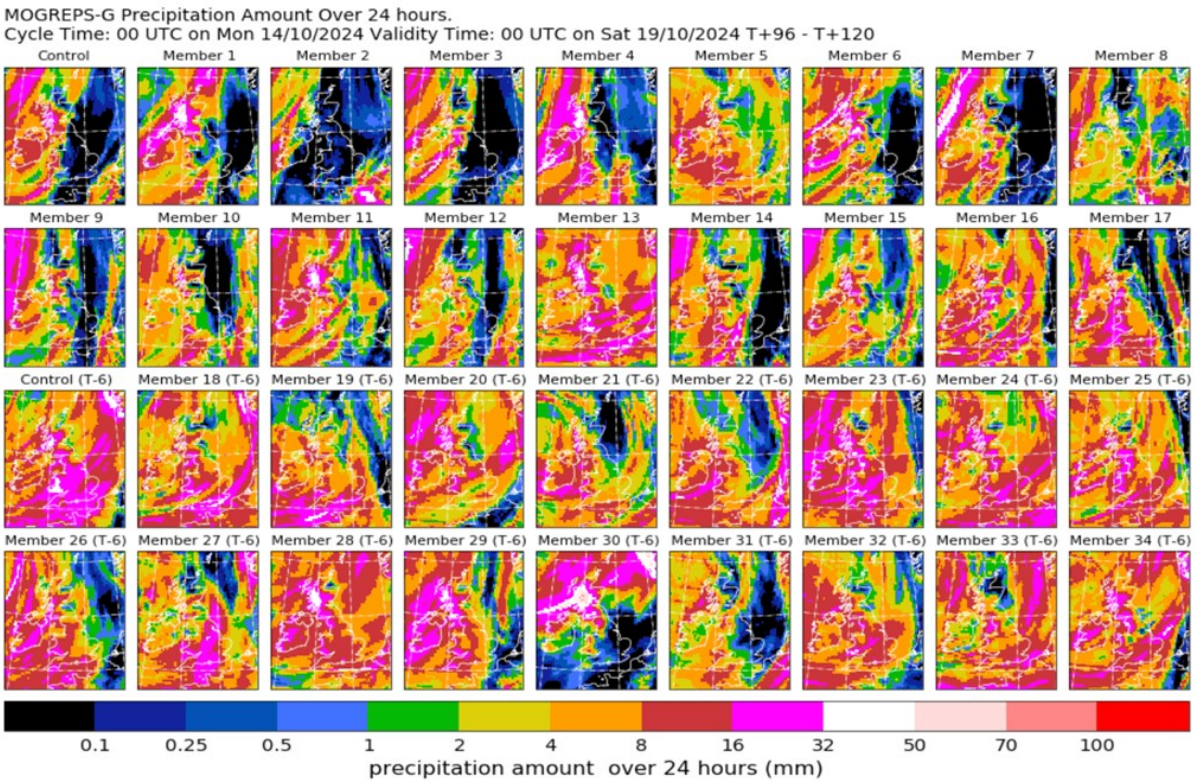


Figure 4.2. MOGREPS-G 24-hour rainfall accumulations valid 00 UTC Friday–00 UTC Saturday (T+96–T+120).

In this case study, one of the two clustering configurations using 6-hour precipitation accumulations as features are going to be the most appropriate. The first of these defines features as anything above the 90th percentile (based on data over all grid points, members and lead times within the same MOGREPS-G run), and the second of these defines features as anything greater than 10 mm in 6 hours. In this case study, features defined using the 90th percentile (which

© Crown copyright 2024, Met Office

was 3.5mm in this run), was deemed to be most appropriate as the 10mm threshold was too high to identify enough grid points as features (initial analysis of the prototype display has suggested that the FSS and resultant clusters work best when between 5 and 10% of the clustering domain is identified as containing a feature).

Several output diagnostics are available to help users assess the relative success of clustering from the latest run i.e. how good are the clusters from today's run relative to normal? "Normal" here is a MOGREPS-G clustering forecast climatology which has been building up since the prototype trial began in July 2024. An example of three of these output diagnostics (cluster similarity, ensemble spread, and neighbourhood size) are shown in Figure 4.3 and explained in more detail below:

The ***cluster similarity*** measure shown in Figure 4.3(1) helps assess how distinct the clusters are relative to just having one cluster. The score ranges between 0 and 1, with lower scores being better. A score of zero would mean that clusters provide maximum distinction compared to having one cluster. A score of one would mean that clusters provide no benefit over having one cluster. Scores will typically be close to one at short lead times when there is very little ensemble spread, and gradually lower (reveal improved segregation or distinctiveness of clusters) as the lead time and ensemble spread increases. In this example, we are interested in window of interest 9 (T+96-T+120), where the green shading shows that clusters are slightly more distinct than normal.

The ***ensemble spread*** shown in Figure 4.3(2) is calculated as the average distance between members irrespective of clustering. Specifically, it is based on the average distance between all member pairs, using the FSS distances shown in the matrices (e.g. as in Figure 4.4) and used by the k-medoids clustering. In this example, the red shading for Window 9 shows the ensemble spread to be slightly less than normal. This is interesting, as although ensemble spread is lower than normal, the clusters are still relatively distinct.

The ***neighbourhood size*** shown in Figure 4.3(3) is used by the FSS. The nature of the FSS means that larger neighbourhoods are required to achieve a good FSS score if the fields being compared are spatially far apart. Therefore, a balance is achieved by finding the smallest neighbourhood size that results in members being more similar than dissimilar at the grid-point level. This balance is achieved when the FSS is ≥ 0.5 (closer to 0.5 the better). The neighbourhood size (number of grid points) associated with a FSS close to 0.5 is shown here. The use of larger neighbourhood sizes compared to normal suggests that members are spatially less close to each other than normal, which will also be closely related to ensemble spread – this is why ensemble spread and neighbourhood size are very similar in the example shown in Figure 4.3.

© Crown copyright 2024, Met Office

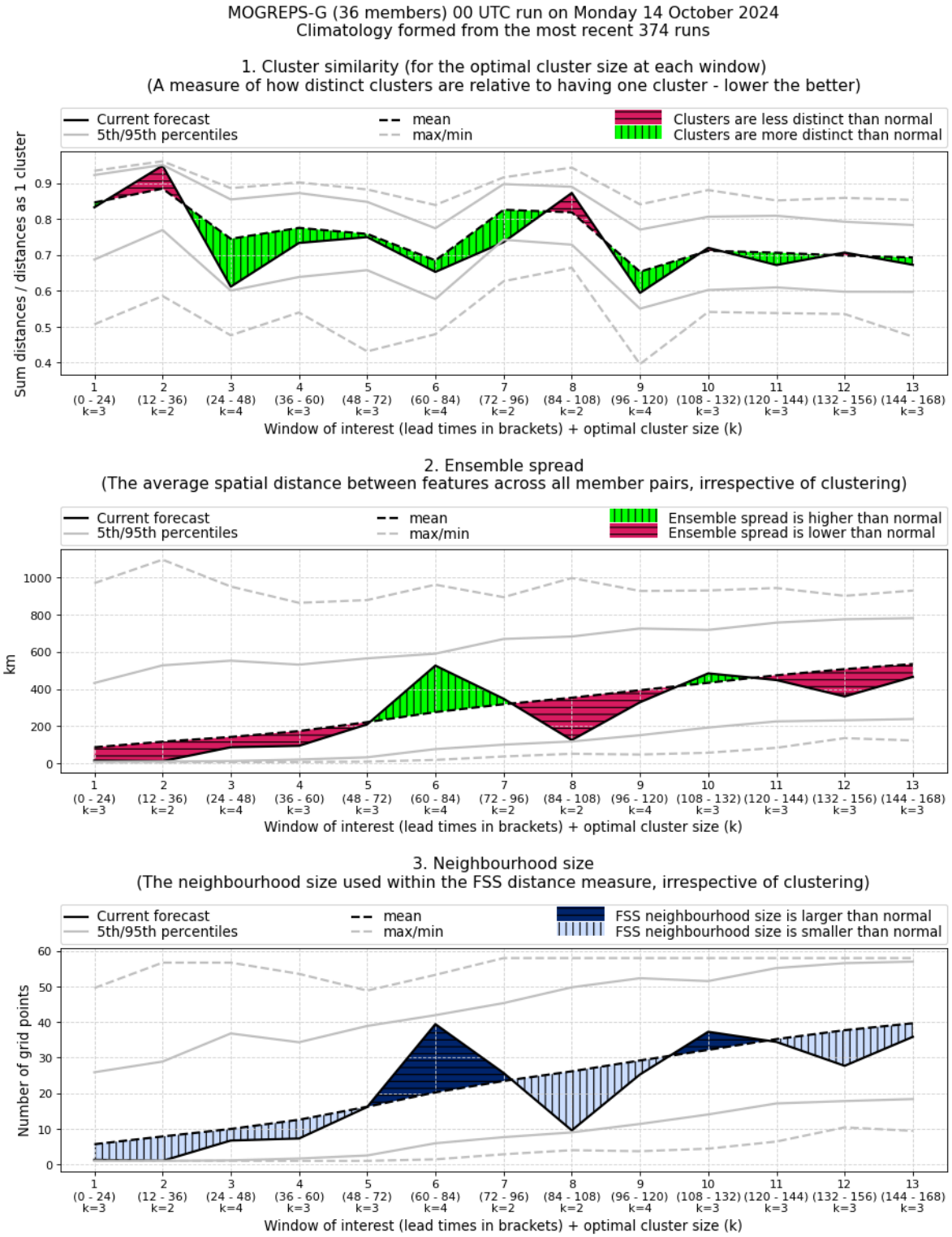


Figure 4.3. Cluster similarity, ensemble spread, and neighbour size metrics, as used for assessing how good clusters are for the latest run.

© Crown copyright 2024, Met Office

Another useful way of assessing how good the current set of clusters are is to look at the distance matrices for the relevant window of interest (e.g. Figure 4.4 for this example case). Here, the unsorted FSS distances are shown on the left and the clustered FSS distances are shown on the right. The yellow boxes highlight the different clusters, which reduce in probability from cluster 1 (bottom left) to cluster 4 (top right). The single dot in each cluster highlights the representative member. These are based on the cluster medoids (central members). The lighter grey shading within each of the yellow boxes suggests the members within are relatively similar. Interestingly, in this example it also appears that clusters 2 and 3 are closely related because an outer light grey box can be seen encompassing them.

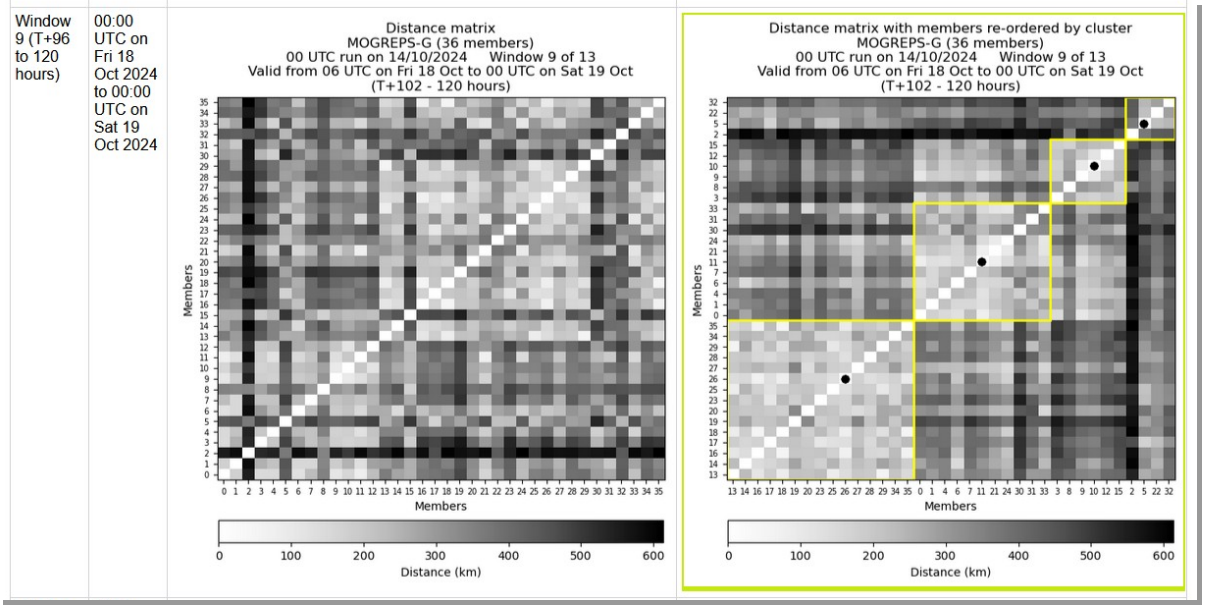


Figure 4.4. Distance matrices for Window 9 (valid on Friday 18th October 2024), showing unordered (left) and order by cluster (right).

By assessing the clustering products so far, we have determined that (1) Window 9 best captures our rainfall event, (2) the optimal number of clusters is 4, but clusters 2 and 3 may be closely related, (3) ensemble spread is slightly less than average (in relation to our features of interest), however this is not an issue because (4) clusters are slightly more distinct than normal. This background information will be useful in interpretating the spatial precipitation patterns seen within each of the four clusters.

© Crown copyright 2024, Met Office

The first place to look for assessing the spatial patterns for each cluster is the aggregated products for each window of interest, which show a quick summary of the clusters for the window in question. The products for window 9 are shown in Figure 4.5:

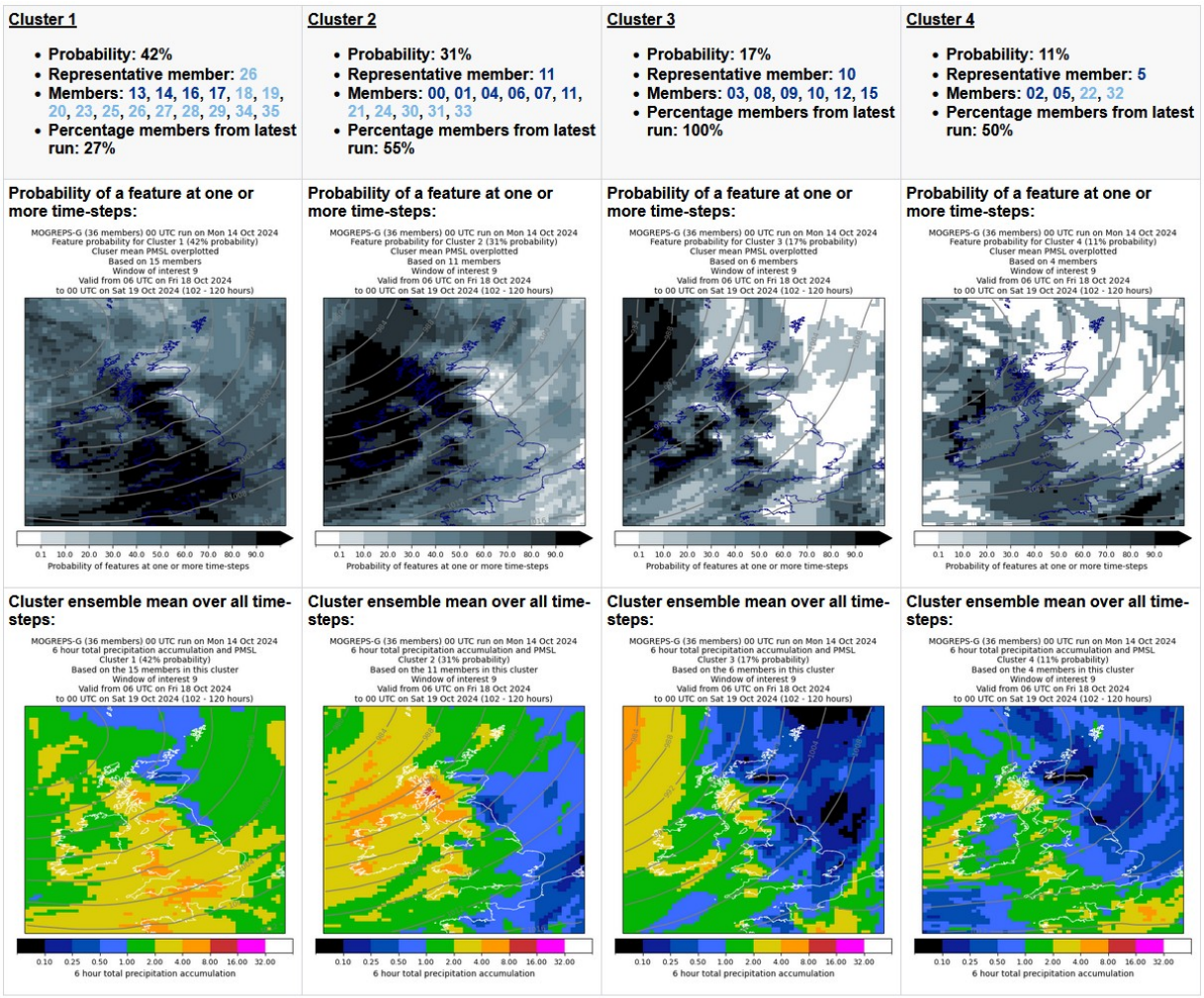


Figure 4.5. Screenshot of Cluster output summary for Window 9 (valid on Friday 18th October 2024), showing cluster probabilities and lagged member membership (top row), cluster feature probabilities (middle row), and cluster means (bottom row).

- The first row (Figure 4.5) gives the ***cluster probabilities*** along with a list of members in each cluster. It is also possible to determine the support each cluster has from members in both the latest and previous 18-member MOGREPS-G run. For example, cluster 3 only contains members from the latest run and cluster 1 is dominated by members from the previous run. This information may be useful to know in rapidly evolving situations when the model is still latching on to new signals that are then affecting the forecast evolution.
- The middle row (Figure 4.5) shows the spatial ***probability of a feature***

© Crown copyright 2024, Met Office

within each cluster. These probability maps are useful for interpreting spread within each cluster and for assessing the focal point of features within each cluster.

- They can be used in conjunction with the **cluster *ensemble mean maps*** (bottom row in Figure 4.5), which provide a mean over all members and time-steps within the given window of interest.

In this example, Cluster 1 has rainfall affecting most western and southern areas of the UK; Cluster 2 has rainfall affecting most western and northern areas of the UK; Cluster 3 is similar to Cluster 2 but with rainfall restricted to more western areas; and Cluster 4 has lighter rainfall in general which is restricted to western coastal stretches. These plots don't provide any information on rainfall timing throughout the day, nor do they visualise the representative members – both of these visualisations and others can be viewed by clicking on the webpage link titled "Detailed forecast information for Window 9" (shown in the screenshot at the bottom of Figure 4.5). This loads a separate page containing cluster information (means and representative members) for each time-step within the chosen window of interest. For example, the time-step (6-hour rainfall accumulation) valid between 0900–1500 hrs on Friday 18th October 2024 is shown in Figure 4.6. This period was chosen as it represents the clusters during the middle part of the day when people may be out and about and therefore more impacted by the weather conditions. Cluster 1 RM shows heavy rainfall in Wales and western/central parts of England, the Cluster 2 RM shows heavy rainfall making it into western Scotland, the Cluster 3 RM shows lighter rain arriving into western areas only, and the Cluster 4 RM shows rainfall in parts of Scotland, as well as a band of rain in SE England (Figure 4.6 (bottom row)). These are all distinct solutions and sit within the envelope of their respective cluster means (Figure 4.6 (top row)).

© Crown copyright 2024, Met Office

Cluster means

Cluster RMs

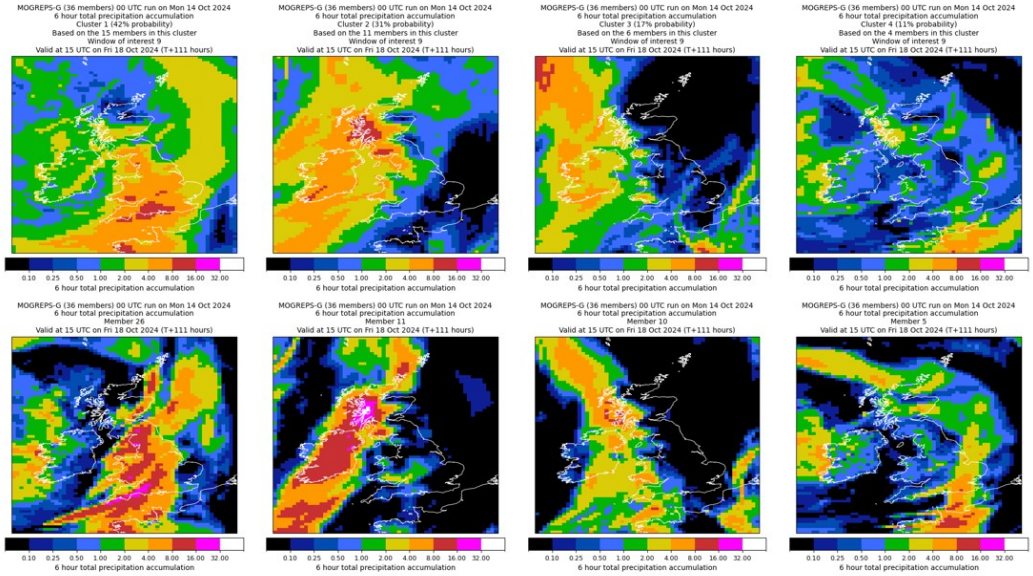


Figure 4.6. Cluster means (top row) and RMs (bottom row) centred around midday on Friday 18th October 2024 (valid between 09:00 and 15:00). Cluster 1 is the first column and Cluster 4 is the last column.

© Crown copyright 2024, Met Office

An example of how these four solutions could be presented to the end user (customer) is shown in Figure 4.7. Here, probabilities have been rounded to the nearest 10% to avoid giving an unrealistic sense of precision with the probabilities, and the scenarios may be suitable for animating over the window of interest, such as within a video presentation.

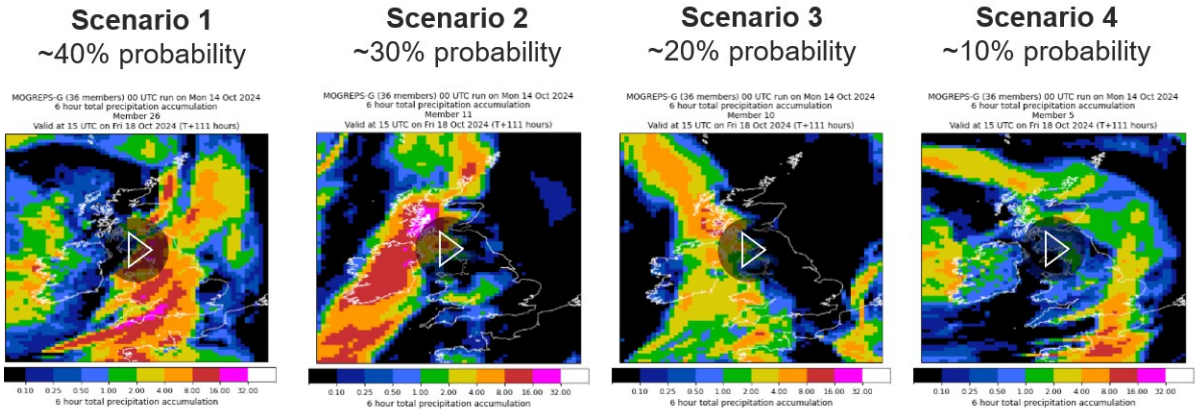


Figure 4.7. Example use of the clustering output to communicate the different forecast scenarios, along with their probabilities. Forecast scenarios could be animated over the window of interest to show how they evolve differently, or they could be presented for a particular time-step of interest, such as one covering the middle part of the day.

4.4. Recommendations for future work

- The feature-based clustering research prototype should continue to be run until the end of March 2025 to allow more case studies to be captured and further feedback to be collected from Op Mets.
- Further scientific research and development work is required prior to any potential operational development of the feature-based clustering, as laid out below.
    1. There are two key recommendations linked to the logic for identifying the optimal number of clusters in each window of interest:
        i. Modify the RM similarity check (shown in Figure 4.1) to use the same absolute distance threshold for each window of interest. This should reduce the number of clusters at earlier lead times, which often look very similar due to a relative threshold being used.
        ii. The minimum number of clusters currently defaults to two. However, there should be an option to have only one cluster in situations when no distinct RMs can be found.
    2. Explore feature-based clustering potential with other global ensembles, e.g. ECMWF ENS, or trialling it on convection permitting ensembles such as MOGREPS-UK. This would allow us to explore how the clustering works on finer scale, when features may be smaller in nature and less spatially coherent. This work idea is being taken forward by Adam Gainford, a PhD

© Crown copyright 2024, Met Office

student at the University of Reading, and will hopefully inform us how best to set up the clustering configurations for MOGREPS-UK in the future. The use of sub-km scale ensembles is also a rapidly evolving area of research (Bain et al., 2022; Hanley and Lean, 2024) and it may be useful to explore the use of clustering on forecast fields of this very high resolution too, in order to aid its interpretation by Op Mets. In addition, the generation of sub-km scale ensembles is computationally very expensive, and so the clustering methods developed here could potentially be adapted to identify a reduced set of distinct members (RMs) to then be run at the sub-km resolution.

3. Experiment with modified distance measures used within the k-medoids clustering. Adam Gainford also hopes to run some experiments with an updated distance measure known as the overlap-adjusted FSS (Skok and Roberts, 2018) and explore its impact on the final clusters.

- Reduce the number of configurations run routinely based on user feedback. There are currently 18 configurations being trialled, split across four parameters (850 hPa WBPT, 6-hour precipitation accumulations, 10m wind gusts, and PMSL). Feedback from Op Mets already suggests little benefit in the PMSL clustering, which could be switched off int he future. In addition, configurations for the other three parameters could be reduced to one or two configurations each, making it easier for users to find what they are looking for. Op Mets have also requested additional clustering output for snowfall accumulations, which would be beneficial in winter in communicating high-impact snowfall scenarios to customers.
- Feedback from Op Mets has also shown that it would be helpful to re-design the user journey through the range of clustering outputs, as several have commented that they find the current layout of the webpages challenging to use, which is reducing their uptake.
- Op Mets would like an on-demand version of the featured based clustering tool, with the user defining inputs such as the domain, parameter, threshold for identifying features, window length and overlap. This would (1) allow the clustering to work in a similar way to ESA on demand tool, and (2) ensure that the resultant clusters are always relevant to the weather situation at the time, specifically with regards to feature identification.
- If the feature-based clustering is developed operationally, one potential option is for the main output of the clustering code to be simple CSV or Json files of cluster membership and RMs, with their visualisation handled by Vortex (the future of visualisation, production and dissemination tooling for the Met Office).Link in with Ensemble Sensitivity Analysis (ESA) and Ensemble Sub-Setting (ESS). Clustering, ESA, and ESS can be used individually, but they could also potentially be used together, as illustrated in Figure 4.8. ESA can be used to understand how different forecast pathways lead to different forecast outcomes (see section 2.1). Results from this may help us understand why

© Crown copyright 2024, Met Office

ensembles are clustering in a particular way at a lead time of interest. ESS can then be used to identify a smaller set of members that have lower errors in early forecasts (see section 2.2). Downstream forecasting products (such as clustering) could then be adjusted or re-run based on this preferred subset of members.

- Finally, explore the use of feature-based clustering for identifying RMs to drive downstream applications, where it is computationally expensive and/or time consuming to run the applications off a full ensemble. Such applications of this may include hazard-impacts models, such hydrological modelling used in flood predictions.

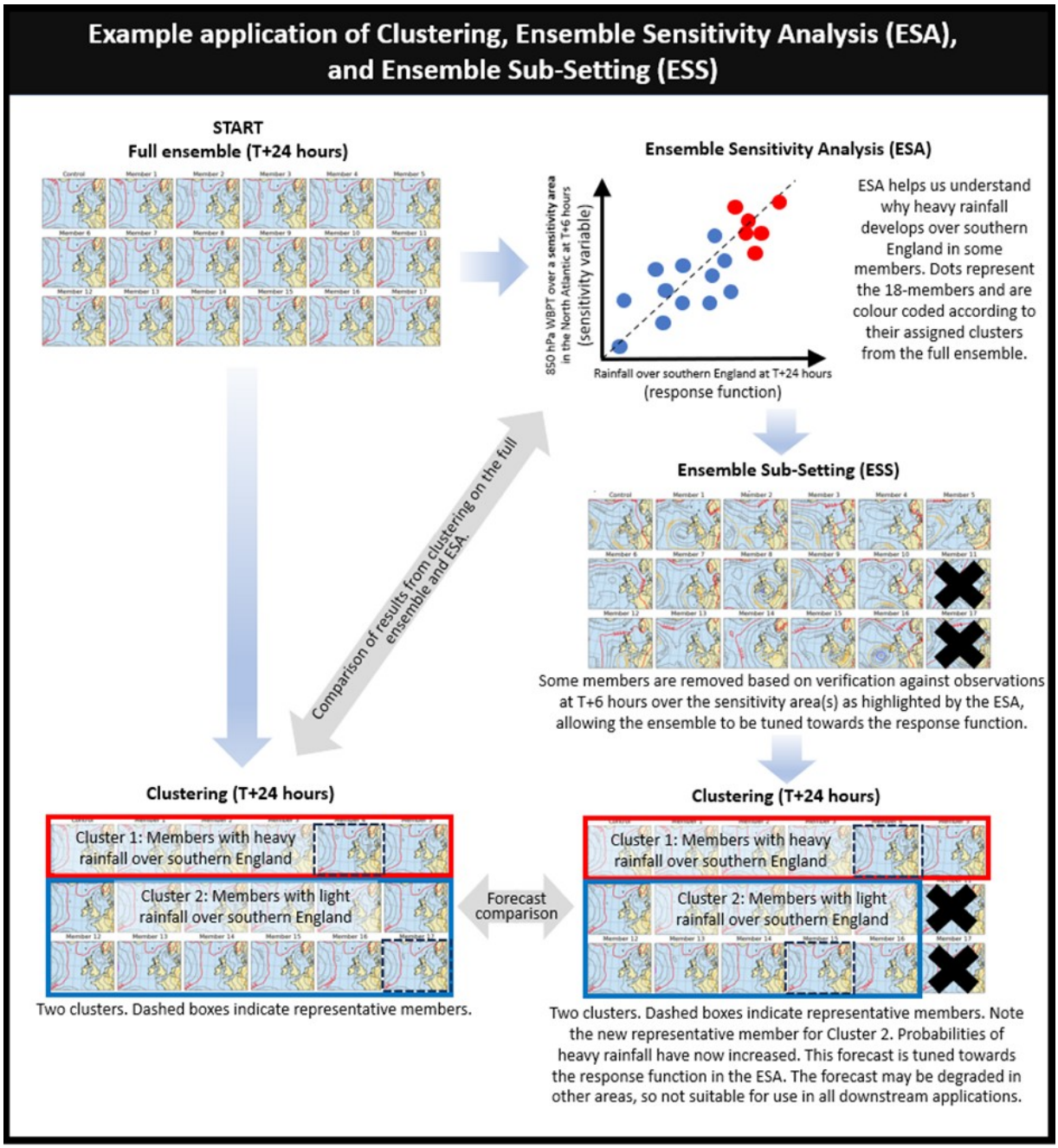


Figure 4.8. A schematic illustrating how clustering, ESA and ESS can be used together to help get

© Crown copyright 2024, Met Office

more information from the ensemble.

© Crown copyright 2024, Met Office

References

Ancell, B. C. (2016): Improving High-Impact Forecasts through Sensitivity-Based Ensemble Subsets: Demonstration and Initial Tests. Weather and Forecasting, 31, 1019–1036, https://doi.org/10.1175/WAF-D-15-0121.1

Ancell, B. C., & Coleman, A. A. (2022). New Perspectives on Ensemble Sensitivity Analysis with Applications to a Climatology of Severe Convection. Bulletin of the American Meteorological Society, 103(2), E511-E530. doi:10.1175/BAMS-D-20-0321.1

Ancell, B., & Hakim, G. J. (2007). Comparing Adjoint- and Ensemble-Sensitivity Analysis with Applications to Observation Targeting. Monthly Weather Review, 135(12), 41174134. doi:10.1175/2007MWR1904.1

Bain, C.L., Bush, M., Buxmann, J.C., Davies, P., Finnenkoetter, A., Flack, D.L.A. et al. (2022) The Met Office winter testbed 2020/2021: experimenting with an on-demand 300-m ensemble in a real-time environment. Meteorological Applications, 29, e2096. Available from: https://doi.org/10.1002/met.2096

Brusco, M., Steinley, D., and Stevens, J. (2019). K-medoids inverse regression. Commun. Stat. Theory, 48:4999–5011

Boykin, K.A. (2022) Extracting likely scenarios from Ensemble Forecasts in real time. PhD in Atmosphere, Oceans and Climate, Department of Meteorology, University of Reading, DOI: 10.48683/1926.00111270

Coleman, A., & Ancell, B. (2020). Toward the Improvement of High-Impact Probabilistic Forecasts with a Sensitivity-Based Convective-Scale Ensemble Subsetting Technique. Monthly Weather Review, 148(12), 4995–5014. doi:10.1175/MWR-D-200043.1

Etheridge, D. (2024). Ensemble Sensitivity Analysis (ESA) Project Report. Internal Met Office Project Report. Available internally at ESA Project Report July 2024.pdf).

Ferranti, L., Corti, S. (2011) New clustering products. ECMWF Newsl. 127: 6–11.

Ferranti, L., Corti, S., Janousek, M. (2015) Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. Quarterly Journal of the Royal Meteorological Society, 141, 916–924.

Gilleland, E., G. Skok, B. G. Brown, B. Casati, M. Dorninger, M. P. Mittermaier, N. Roberts, and L. J. Wilson, 2020: A Novel Set of Geometric Verification Test Fields with Application to Distance Measures. Mon. Wea. Rev., 148, 1653–

© Crown copyright 2024, Met Office

1673, https://doi.org/10.1175/MWR-D-19-0256.1

Hanley, K.E., Lean, H.W. (2024) The performance of a variable-resolution 300-m ensemble for forecasting convection over London. Quarterly Journal of the Royal Meteorological Society. 763, 3737-3756.

Lamberson, W.S., Bodner, M.J., Nelson, J.A., Sienkiewicz, S.A. (2023) The Use of Ensemble Clustering on a Multimodel Ensemble for Medium-Range Forecasting at the Weather Prediction Center. Weather and Forecasting, 38: 539-554, https://doi.org/10.1175/WAF-D-22-0154.1

Mounier, A. Raynaud, L., Rottner, L., Plu, M. (2022) Rainfall scenarios from AROME-EPS forecasts using autoencoder and climatological patterns. Presented at the ECMWF Machine Learning workshop on 31/03/2022. https://events.ecmwf.int/event/294/contributions/3059/attachments/1758/3181/ML-WS_Mounier.pdf, Accessed 27/11/2023.

Mounier, A., Raynaud, L., Rottner, L., Plu, M. (2023) Bow echo detection and rainfall scenarios: two ways to extract relevant information from convective-scale ensembles, EMS Annual Meeting 2023, Bratislava, Slovakia, 4–8 Sep 2023, EMS2023-84, https://doi.org/10.5194/ems2023-84, 2023.

Neal, R., Fereday, D., Crocker, R., Comer, R.E. (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. Meteorological Applications, 23, 389-400.

Neal, R., Robbins, J., Crocker, R., Cox, D., Fenwick, K., Millard, J., Kelly, J. (2024) A seamless blended multi-model ensemble approach to probabilistic medium-range weather pattern forecasts over the UK. Meteorological Applications, 31, e2179. https://doi.org/10.1002/met.2179

Omran, M., Engelbrecht, A., and Salman, A. (2007). An overview of clustering methods. Intell. Data Anal., 11:583–605.

Porson, A , Ackerley, D., Almond, C., Ayliffe, B., Boykin, K., Bulmer, C., Claxton, B., Derbyshire, S., Edwards, J., Evans, G., Gallagher, S., Hanley, K., Hattersley, E., Keates, S., Lewis, H., Lyth, D., Mahmood, S., McCabe, A., Moseley, S., Mylne, K., Oliver, D., Roberts, N., Thornhill, A., Rosling, W., Rust, F., Stone, E., Walker, B., Walters, D., Willington, S. (2022) Winter 2022 Testbed. Met Office internal report.

Roberts, N., Lean, H.W. (2008) Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. Monthly Weather Review, 136: 78-97

Skok, G., Roberts, N. (2018) Estimating the displacement in precipitation forecasts using the Fractions Skill Score. Quarterly Journal of the Royal Meteorological Society, 133: 414-425

© Crown copyright 2024, Met Office

Torn, R. D., & Hakim, G. J. (2008). Ensemble-Based Sensitivity Analysis. Monthly Weather Review, 136(2), 663-677. doi:10.1175/2007MWR2132.1

Met Office
FitzRoy Road
Exeter
Devon
EX1 3PB
United Kingdom

© Crown copyright 2024, Met Office